EXAMINATION OF PSYCHOMETRIC PROPERTIES OF A TRANSLATED SOCIAL-EMOTIONAL SCREENING TEST: THE TAIWANESE VERSION OF THE AGES AND STAGES QUESTIONNAIRES: SOCIAL-EMOTIONAL

by

CHIEH-YU CHEN

A DISSERTATION

Presented to the Department of Special Education and Clinical Science
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

March 2017

DISSERTATION APPROVAL PAGE

Student: Chieh-Yu Chen

Title: Examination of Psychometric Properties of a Translated Social-Emotional Screening Test: The Taiwanese Version of the Ages and Stages Questionnaires: Social-Emotional

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Special Education and Clinical Sciences by:

Dr. Jane Squires	Chairperson
Dr. Debra Eisert	Core Member
Dr. Jeffrey Sprague	Core Member
Dr. Kathleen Scalise	Institutional Representative

and

Scott L. Pratt	Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded March 2017

DISSERTATION ABSTRACT

Chieh-Yu Chen

Doctor of Philosophy

Department of Special Education and Clinical Sciences

March 2017

Title: Examination of Psychometric Properties of a Translated Social-Emotional Screening Test: The Taiwanese Version of the Ages and Stages Questionnaires: Social-Emotional

Investigating the psychometric properties of a screening instrument for young children is necessary to ascertain its quality and accuracy. In light of the important role culture plays on human beliefs and parenting styles, a newly translated and adapted test needs to be studied. Evaluating outcomes on a translated version of a test may reveal significant information related to cultural specifications as well as the common nature of child development.

The current study examined psychometric properties of the 48-month interval of the Ages and Stages Questionnaires: Social-Emotional Second Edition (ASQ:SE-2) and its Traditional Chinese version (ASQ:SE-TC), using item response theory (IRT). Participants in the U.S. included 3,005 young children/parents dyads; 1,455 dyads were collected to represent a Taiwanese sample.

A two-dimensional Rasch Partial Credit Model (2D-RPCM), which was determined to present a better fit than a unidimensional Rasch Partial Credit Model, was used to examine the item fit, item difficulty, reliability, and item information curves to evaluate the psychometric properties on the ASQ:SE and ASQ:SE-TC. Further, differential item functioning was conducted to examine whether items were functioning

differently in the two population groups. Lastly, the differences between the distributions of children's latent traits on the continuum of social and emotional competencies for the U.S. and Taiwanese samples were investigated.

Based on findings, the adequacy of psychometric properties is discussed, providing insight into the quality of particular items. Identified differences between the two populations are explored by reviewing literature regarding cultural comparisons of childrearing practices, parenting styles, and cultural beliefs. Future directions for research include examining the cultural equivalence between translated and original versions of other ASQ:SE-2 intervals.

CURRICULUM VITAE

NAME OF AUTHOR: Chieh-Yu Chen

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro
Taipei Municipal University of Education, Taipei
National Taichung Teachers College, Taichung

DEGREES AWARDED:

Doctor of Philosophy, Special Education, 2017, University of Oregon
Master of Arts, Special Education, 2006, Taipei Municipal University of Education
Bachelor of Education, Special Education, 2004, National Taichung Teachers College

AREAS OF SPECIAL INTEREST:

Child Development
Psychometrics

PROFESSIONAL EXPERIENCE:

Graduate Teaching Faculty, University Center for Excellence in Developmental Disabilities, University of Oregon, 2015-2016

Research Assistant, Department of Special Education, Taipei Municipal University of Education, 2011-2012

GRANTS, AWARDS, AND HONORS:

Doctoral Research Award, College of Education, University of Oregon, 2016

Davis-Bricker Scholarship, College of Education, University of Oregon, 2015

Doutorado-Sanduiche Reverso Fellowship, The State of Rio de Janeiro Research Foundation, Brazil, 2014-2015

Jim and Jean Scholarship, International Student Affairs, University of Oregon, 2013

Studying Abroad Scholarship, Ministry of Education, Taiwan, 2012-2014

Graduate School Scholarship, Taipei Municipal University of Education, 2005-2006

PUBLICATIONS:

Chen, C. Y., Xie, H. C., Chen, C-I., Clifford, J., Squires, J. (accepted). Examining internal structures of a developmental measure using multidimensional item response theory. *Journal of Early Intervention*.

Chen, C. Y., Xie, H. C., Squires, J., Filgueiras, A., Anunciação, L., Landeira-Fernandez, J. (accepted). Examining the psychometric properties of the Brazilian Ages & Stages Questionnaires-Social-Emotional: Use in public child daycare centers in Brazil. *Journal of Child and Family Studies*.

Chen, C. Y., Chen, C-I., Xie, H. C., Squires, J. (in press). Using differential item functioning to examine the gender difference in social-emotional competence of infants and young children. *Bulletin in Early Intervention*. (text in Chinese).

Chen, C. Y., Chen, C-I., Squires, J., Bian, X. Y., Heo, K. H., Filgueiras, A., Kalinina, S., Samarina, L., Ermolayeva, E., Xie, H. C., Yu, Y. T., Wu, P. F., Landeira-Fernandez, J. (2017). Adapting a developmental screening measure: Exploring the effects of language and culture on a parent-completed social-emotional screening test. *Infants and Young Children*, *30*(2), 111-123.

Bian, X. Y., Xie, H. C., Squires, J., Chen, C. Y. (2017). Adapting a parent-completed, social-emotional questionnaire in China: The Ages & Stages Questionnaires: Social-Emotional. *Infant Mental Health Journal*. Advance online publication.

Xie, H. C., Chen, C-I., Chen, C. Y., Squires, J., Li. W., Zhang. Y. (2016). Developing a home-based early intervention personnel training program in Southeast China. *Topics in Early Childhood Special Education*. Advance online publication.

Chen, C. Y., Filgueiras A., Squires, J., Landeira-Fernandez, J. (2016). Examining the factor structure of an early childhood social emotional screening assessment. *Journal of Special Education and Rehabilitation*, *17*(3-4), 89-104.

Chen, C. Y., Squires, J., Heo, K. H., Bian, X. Y., Chen, C-I., Filgueiras, A., Xie, H. C., Murphy, K., Dolata, J., & Landeira-Fernandez, J. (2015). Cross cultural gender differences in social-emotional competence of young children: Comparisons with Brazil, China, South Korea, and the United States. *Mental Health in Family Medicine*, *11*(2), 59-68.

Acar, S., Hansen, S., Dolota, J. & Chen, C. Y. (2014). An overview of developmental screening: Implications for practice. *Baskent University Journal of Education, 1*(1), 9-18.

Chen, C-I., & Chen, C. Y. (2013). A reflection on the development of developmental screening instruments for young children: Using the Ages and Stages Questionnaires research team in the United States as a model. *Special Education for the Elementary School*, *55*, 27-41. (text in Chinese).

Tsai, K. Y., & Chen, C. Y. (2011). The development and assessment of social emotional behavior of infants and young children. *Special Education for the Elementary School*, *51*, 29-40. (text in Chinese).

Chen, C. Y. (2010). Exploring the feasibility of using modern board games in children's learning activities. *The Elementary Education Journal*, *57*(4), 40-45. (text in Chinese).

Chen, C. Y., & Tsai, K. Y. (2009). The development of the emotional and behavioral problems checklist for young children. *Psychological Testing, 56*(2), 235-268. (text in Chinese).

Tsai, K. Y., & Chen, C. Y. (2009). *A preliminary study on the development of the emotional and behavioral problems checklist for young children.* Taipei: Special Education Center of Taipei Municipal University of Education.

Chen, C. Y. (2008). Investigating the issues of early intervention assessment by the transactional model. *Special Education for the Elementary School*, *45*, 78-83. (text in Chinese).

Chen, C. Y. (2005). Using trans-disciplinary play-based assessment to young children with developmental delay. *Special Education for the Elementary School, 40*, 35-43. (text in Chinese).

ACKNOWLEDGMENTS

Many thanks to all the people who have supported me during these years with your knowledge, your caring and your board games from all walks of life on Earth. I will do my best in my field to bring benefit to this world to return your favor.

TABLE OF CONTENTS

Chapter Page

Chapter Page

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

The critical nature of social-emotional competence in young children has been emphasized in recent years. Prevalence studies suggest that 13% to 25% of young children have at least mild emotional and/or behavioral problems (Campbell, 1995; Jellinek et al., 1999; Lavigne et al., 1996). Growing evidence indicates that social-emotional problems in young children can have long-term effects (Feil, Walker, & Severson, 1995; Lavigne et al., 1998; Shaw, Gilliom, Ingoldsby, & Nagin, 2003); problems that go untreated can become serious and costly to treat as children age (Bricker, Davis, & Squires, 2004; Slonim, 1991). In addition, social-emotional competence is crucial for school readiness for young children and serves as an essential foundation for learning skills that are important to academic performance in later years (Denham, 2006; Jones & Bouffard, 2012; Zins, Bloodworth, Weissberg, & Walberg, 2007). Given the importance of social-emotional competence, using technically sound screening instruments for the early identification of social-emotional problems is critical, as well as providing timely intervention once problems are detected (Bricker et al., 2004; Briggs-Gowan, Carter, Irwin, Wachtel, & Cicchetti, 2004; Briggs et al., 2012). Screening tests with solid technical adequacy help identify children who may be at risk for social-emotional delays and assist in early identification and referral for needed services. Furthermore, early identification leads to interventions that may improve the development of children with developmental delays, and reduces the expenditures for remedial services, promoting children's long-term success (Garces, Thomas, & Currie, 2000; Reynolds, Temple, Robertson, & Mann, 2001, 2002). Thus, investigating the

psychometric properties of screening tools is necessary to assure accuracy, especially when tests are translated and used across cultures.

Human development has been thought of as an interactive process, with biological and cultural factors interacting (Cole, 1998; Keller, 2007). Culture strongly affects human attitudes, beliefs, values and behaviors (Slonim, 1991), including parents' perceptions of appropriate social-emotional performance (Hetherington, Parke, & Locke, 1999). Defining an appropriate social-emotional behavior is complicated because it depends on and is affected by cultural and family values, developmental level, and situational circumstances (Bricker & Squires, 2014). Parents from differing cultural backgrounds transmit their own beliefs, values, and attitudes to their children and attempt to shape their children as successful individuals according to their cultural expectations (Harkness & Super, 1996; Okagaki & Sternberg, 1993; Tudge, 1991; Vinden, 2001). An appropriate behavior in one culture for children might be discouraged in others (Gauvain, 2001). In light of the different perceptions of behavior, researchers have advocated that cultural influences should be emphasized in developmental assessments of young children (Betancourt & Lopez, 1993; Coll & Magnuson, 2000).

In early intervention, there are four main purposes for assessment (Neisworth & Bagnato, 2004): (1) screening and eligibility determination, (2) individualized program planning, (3) children progress monitoring, (4) and program evaluation. Screening is the initial step to identify children who need more in-depth assessment of developmental needs. Conducting a screening procedure for infants and young children with developmental concerns usually relies on parent/caregiver report and experts recommend using reliable, standardized instruments (Slentz, Early, & McKenna, 2008). Screening

tests help identify children with social-emotional delays, reduce the cost of assessment, and assist in referral for needed services (Neisworth & Bagnato, 2004).

The Ages & Stages Questionnaires: Social-Emotional (ASQ:SE) is a series of parent completed screening questionnaires for 1-to-72 month old children, designed to identify potential social-emotional delays as soon as they occur (Squires, Bricker, & Twombly, 2002; Squires, Bricker, & Twombly, 2015). With adequate psychometric properties for screening social-emotional problems (Briggs-Gowan et al., 2004; Briggs et al., 2012; Carter, Briggs-Gowan, & Davis, 2004; Marks & LaRosa, 2012), the ASQ:SE has been widely used in Head Start and Early Head Start programs across the United States (Baggett, Warlen, Hamilton, Roberts, & Staker, 2007; Beeber et al., 2010) as well as internationally (Bian, Wang, & Chen, 2013; Heo & Squires, 2012). Since the ASQ:SE has been used internationally over the last decade, it is important to thoroughly re-examine its psychometric properties, in order to assure effectiveness. Psychometric research can improve the practice of early identification and benefit subsequent assessment and intervention practices.

The purpose of this research is to examine the psychometric properties in one widely-used social emotional screening test in the U.S., the ASQ:SE Second Edition, and to compare the original and Taiwanese translated versions using item response theory (IRT). Conducting IRT analyses may help test developers examine the quality of each item and understand how the item functions across the continuum of the targeted latent trait.

Investigating the consistency of item functioning across these two cultures is the second research aim. If item functions are similar across cultures, inherent similarity in

the nature of social-emotional competence and/or underlying cultural attitude might be an explanation. If items function differently across cultures, cultural practices and attitudes may be the reason. With the exploration of psychometric properties of a widely used instrument, and the successive interpretation from cultural viewpoints, the results of this study may contribute to the body of knowledge in child development and the nature of social-emotional competence across cultures.

CHAPTER II

REVIEW OF LITERATURE

Introduction

A review of the literature is presented on topics related to this study: (1) social-emotional competence, (2) assessment of social-emotional competence, (3) effects of culture on social-emotional development (4) the ASQ:SE related research, and (5) item response theory.

Social-Emotional Competence

Before the implementation of PL 94-142 and PL 99-457, which addressed issues related to services and evaluation in social and emotional developmental areas for children, researchers and policy makers had pointed out the critical need for early identification of young children's social-emotional problems (Collins, 2002). The demand for early intervention and services focusing on young children's social-emotional competence increased after the federal legislation in the 1970s and 1980s (Trohanis, 2008). Researchers have suggested that 37% of 18-month-olds who demonstrate social-emotional problems will carry these problems into their preschool years (Mathiesen & Sanson, 2000). For young children with severe problem behaviors, it was estimated that more than 50% continue to have problem behaviors 1 to 2 years later (Lavigne et al., 1998). Researchers noted that the quality of children's social-emotional competence is related to success in early adjustment to school, better relationships with peers and teachers, and higher academic performance (Ladd, Birch, & Buhs, 1999; Ladd, Kochenderfer, & Coleman, 1996). The following section reviews the definitions and parameters of social-emotional competence.

Definition and Content

The terms social-emotional competence, emotional behavioral problems, and mental health often refer to similar underlying characteristics (Briggs-Gowan & Carter, 2008; Briggs-Gowan, Carter, Skuban, & Horwitz, 2001; Mathiesen & Sanson, 2000; Merrell & Holland, 1997; Noll et al., 1999; Squires, Bricker, Heo, & Twombly, 2001; Whitcomb & Merrell, 2013). Social competence and emotional competence are highly related but represent different developmental traits and behavioral processes (Squires et al., 2015). Social competence is considered a series of behaviors that allow one to have positive relationships with others (Jones & Bouffard, 2012; Raver & Zigler, 1997). Emotional competence is the personal underlying ability to regulate one's emotion to accomplish his/her goals (Campos, Mumme, Kermoian, & Campos, 1994).

Specifically, researchers suggested five dimensions for social-emotional development including: (1) social competence; (2) attachment; (3) emotional competence; (4) self-perceived competence; and (5) temperament/personality (Denham, Wyatt, Bassett, Echeverria, & Knox, 2009). First, *social competence* can be defined as the ability to develop appropriate social interactions with people (Denham, 1998). Some representing behaviors of social competence include having interest in people, displaying attention, initiating contact with people, playing alongside friends, participating in playing within a group, and beginning of specific friendships.

Attachment may be defined as an ability that the child develops from the very early years, which enables him or her to build positive connections with parents and caregivers. Attachment forms the foundation for a person to have a close relationship to others. A third dimension, emotional competence can be defined as a complex ability to

realize self and others' emotion, to negotiate interpersonal exchanges, and to control and adjust emotions (Saarni, 1999). Behaviors include expression of social emotions and differential reaction to other's emotions, and emotion regulation. Children who are deficient in social-emotional competence may experience difficulties in social interactions and have problem behaviors (Masten & Coatsworth, 1995).

Fourth, self-perceived competence can be defined as a child's evaluation of his/her own competence including cognitive, physical and social abilities (Denham et al., 2009). The representing behaviors include responding to one's own name, recognizing self, expressing ownership, speaking positively of self, and showing autonomy. Finally, temperament/personality refers to the individual differences in emotional reactivity and self-regulation (Denham et al., 2009).

Assessment of Social-Emotional Competence

Assessment--yielding a developmental profile of a child--is a process of collecting information for the purpose of making decisions, which may be used to determine which individuals need specific services (McLean, Bailey, & Wolery, 1996). Developmental assessments can be varied regarding their purposes (e.g., screening, diagnosis, program planning or monitoring) and can focus on different specific developmental areas (e.g., motor skills, communication, or social-emotional competence). Early identification of social-emotional problems for infants and young children is critical for providing appropriate interventions (Squires, 2000). To identify social-emotional problems early on, accurate instruments are required (Bricker et al., 2004; Briggs et al., 2012). Without using standardized developmental screening instruments to evaluate the developmental status of children, pediatricians' judgment failed to identify 60% to 80% of children with

developmental delays (Halfon et al., 2004; Sand et al., 2005). The following section focuses on topics related to assessment in social-emotional competence in early childhood.

Dimensional and Categorical Measurement

To measure social-emotional competence/behavioral problems, there are two main classifications: dimensional measurement and categorical measurement. Dimensional measurement relies on the rating of presence and severity of various behavioral and/or emotional performances. Behavioral performance data can be clustered into dimensions, using statistical techniques. For example, the Child Behavior Checklist for Ages 1.5-5 (Achenbach & Rescorla, 2000) and Infant-Toddler Social and Emotional Assessment (Carter, Briggs-Gowan, Jones, & Little, 2003) have been analyzed using confirmatory factor analysis on large populations to investigate the latent constructs of emotional and behavioral performance. The advantage of dimensional measurement is that a judgment about a child's behaviors is made from comparing the testing results to a large normative sample, whereas the disadvantage is the inability to determine a detailed diagnosis from the results (Dunn, Austin, & Perkins, 2009). Dimensional measurement usually presents as a questionnaire or a checklist, for example, the ASQ:SE and the Infant-toddler Social and Emotional Assessment (Carter et al., 2003). In contrast, categorical measurement focuses on problems by inspecting if the presence of behavioral or emotional performance reaches a specific criterion, such as the Diagnostic and Statistical Manual of Mental Disorders, (DSM-5) (American Psychiatric Association, 2013) and Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood (Zero To Three, 2005). The results of categorical

measurement can be used to guide treatment, but cannot indicate the severity compared to the population (Dunn et al., 2009).

Difficulty and Challenges

Assessing social-emotional competence and problems in early childhood is a challenging process (DelCarmen-Wiggins & Carter, 2004; Zeanah, Boris, & Larrieu, 1997). Defining an appropriate social-emotional behavior in early childhood is complicated and challenging because children's developmental level, situational circumstances, and cultural/family values affect the interpretation of performance (Squires et al., 2015). Rapid development in early childhood, combined with assessing the child within relational and cultural contexts are challenging (Carter et al., 2004). Neisworth and Bagnato (2004) argued that assessment practices in a natural context must measure the authentic performance of infants and young children in order to be valid. The natural context includes: (1) naturally occurring behaviors, (2) natural environment (e.g., everyday circumstances), and (3) natural observations. Under the philosophy of assessing children in the natural environment, assessment tools such as the ASQ:SE measure daily occurring behaviors in natural settings. Parents/caregivers assess their child at home or in daycare, through direct observation, rather than limiting observations to the clinic or laboratory.

Culture and Social-Emotional Development

The ecological model, an important developmental theory, suggests inspection of human development within cultural contexts (Bronfenbrenner, 1979). This model considers human behavior as a result of two-directional interactions between characteristics of the individual and the environment, along with contemporary and

historical contexts (Gardiner, Mutter, & Kosmitzki, 2008). Parents, based on their beliefs, which are shaped by their own culture, raise their children to become successful members in their culture (Gardiner et al., 2008; Harkness & Super, 1996; Okagaki & Sternberg, 1993; Tudge, 1991; Vinden, 2001). This process can be called "socialization". Socialization varies by culture, which may result in contrasting outcomes in various developmental areas such as motor skills, play pattern, personality, cognition, social behavior and gender roles (Gardiner et al., 2008).

Social-emotional competence is a developmental area strongly influenced by culture. Culture plays an important role in parenting so that the influence starts from early infancy. For example, sleeping arrangements can be very different across cultures (McKenna & Gettler, 2007). Lynch and Hanson (2011) noted that mainstream American culture encourages young children to sleep in separate rooms from their parents or siblings. By contrast, parents from families in other cultures (e.g., families from Asia or South America) prefer that young children sleep with adults. Feeding practices can also vary across cultures. For example, in Chinese cultures, mealtimes are highly structured with food expected on time, whereas some cultures do not have formal preparations and standards. Besides differing family routines, social communication and emotion expression can be different from culture to culture. In mainstream American culture, eye contact is valued, but in other cultures eye contact may be interpreted as impolite or aggressive (Lynch & Hanson, 2011). In another example, Robarchek and Robarchek (1998) found aggression and violence were encouraged in the Waorani of Amazonian Ecuador as opposed to other cultures in similar living conditions. Perez and Gauvain (2007) indicated that cultural beliefs can affect the kinds of play that parents encourage

(i.e., with other peers or with adults) depending on their beliefs about how to achieve developmental goals. For instance, Farver (1999) found that European American mothers consider playing with their children as critical to their social development and that they spend more times playing with their children than mothers in some other cultures. Therefore, to define an appropriate social-emotional behavior is complicated, and cultural belief and family values must be taken into account (Bricker & Squires, 2014).

Ages & Stages Questionnaires: Social-Emotional

The Ages & Stages Questionnaires: Social-Emotional (ASQ:SE) is a screening instrument exclusively focused on a child's social-emotional behaviors. The first edition was published in 2002 (Squires et al., 2002), consisting of a series of eight questionnaires (i.e., 6, 12, 18, 24, 30, 36, 48, and 60 months). The second edition, published in 2015 (ASQ:SE-2; Squires et al., 2015), added a two-month-old interval for detecting social-emotional behavioral problems in very young infants and expanded the age for children from one month to 72 months old. Items on the ASQ:SE address essential social-emotional developmental milestones and problem behaviors. In the second edition, several new items were added to evaluate early social-communication, adaptive, and autonomous behaviors. The current review will investigate the psychometric properties of the ASQ:SE and studies related to the ASQ:SE.

Psychometric Investigations

The initial psychometric study of the ASQ:SE focused on internal consistency, test-retest reliability, determination of cutoff points, and concurrent validity of the 24- and 36-month ASQ:SE (Heo, 1999). With the publication of the ASQ:SE, psychometric properties of eight intervals ranging from the 3- to 63-months were reported, based on

3014 questionnaires (Squires et al., 2001). Concurrent validity was examined using the Infant Toddler Social-Emotional Assessment, the Temperament and Atypical Behavior Scale Screener, and the Child Behavior Checklist, with 90 children/caregivers (Davis, 2002).

The *ASQ:SE Use's Guide* reported high reliability, internal consistency, sensitivity, and specificity (Squires et al., 2002). The effect of the risk/disability status and gender on the performance of the ASQ:SE was examined. Groups labeled as low risk, at risk, developmental disabilities, and social emotional disability were significantly different at all age intervals. Significant differences between genders were found at the 30-, 36-, 48-, and 60-month age intervals (Squires, Bricker, & Twombly, 2004). A comparison study of receiver operating characteristics and item response theory approaches was conducted, and the results underscored the psychometric integrity of the ASQ:SE cutoff scores (Yovanoff & Squires, 2006).

For the psychometric study of the ASQ:SE:2, the developers collected more than 14,000 questionnaires on children from ranging from one to 72 months between 2009-2013, reflecting the U.S. 2010 census data on variables including race/ethnicity, mother's education level, and family income (Squires et al., 2015). Sensitivity ranged from 77% to 84% and specificity and specificity from 76% to 98%. Cronbach coefficient alpha ranged from .71 to .90.

ASQ:SE Screening System Investigations

The Enhancing Developmentally Oriented Primary Care project of the Illinois chapter of the American Academy of Pediatrics and the Illinois Department of Healthcare and Family Services initiated a project to improve the preventive health and

developmental services for children in Illinois (Allen, Berry, Brewster, Chalasani, & Mack, 2010). In this study, the ASQ:SE was found to facilitate identification of children at risk and those who need referral for further evaluation. Gilkerson and Kopel (2005) also noted that using the ASQ:SE increased the screening rate of children with social-emotional development needs in educational and social service settings.

ASQ:SE as an Outcome Measure

Briggs et al. (2012) suggested using the ASQ:SE as a universal screening tool in pediatric settings can identify significant percentages of children who are at risk for social-emotional problems. Additional research using the ASQ:SE in pediatric settings suggested an improvement in primary prevention and early intervention for social-emotional problems (Brown, Copeland, Sucharew, & Kahn, 2012). In a study with Latina mothers, the ASQ:SE was used to measure children's behavioral problems in Early Head Start (Beeber et al., 2010). Jee et al. (2010) noted that using the ASQ-SE increased detection rates for social-emotional problems among young children in foster care. Hillen, Gafson, Drage, and Conlan (2012) focused on identifying the prevalence of developmental delay in England using the ASQ:SE as well as other measures, and the related need for interventions for preschool children. Their results suggested that children's needs and opportunities for early intervention will be improved by using age-appropriate assessments.

ASQ:SE Cultural Studies

The ASQ:SE has been translated into Korean, Simplified Chinese, Traditional Chinese, Russian, Brazilian Portuguese, and Turkish. Psychometric properties were studied based on local populations, and the results including internal consistency, test-

retest reliability, and validity results were adequate (Bian et al., 2013; Heo, Lee, & Squires, 2012; Heo & Squires, 2012; Kucuker, Kapci, & Uslu, 2011). In addition to psychometric studies on cultural/translational versions of the ASQ:SE, gender differences in social-emotional competence were compared using the ASQ:SE across cultures, including samples from Brazil, China, South Korea, and the United States (Chen et al., 2015). Another study investigated the cultural equivalence of the 60-month interval of the ASQ:SE first edition by examining whether the items functioned differently in the original English version compared with five adapted translated versions (i.e., Korean, Portuguese, Russian, Simplified Chinese, and Traditional Chinese), as well as exploring cultural considerations resulting from identified differences (Chen et al., 2017). The result indicated that a large portion of items were identified, suggesting that diverse cultural values, beliefs and expectations across cultures may affect parents' responses.

Item Response Theory

Introduction

Item response theory (IRT) is a statistical theory regarding the relation between an examinee's response measured by test items and an underlying latent trait (Embretson & Reise, 2000; Hambleton & Swaminathan, 1985). An underlying trait, usually labeled by the Greek letter theta (θ), is a proposed ability (e.g., intelligence, emotion regulation) that we expect to measure. Specifically, the essential element in IRT modeling is to establish the relation between theta and the probability of certain response of an item. The relation can be drawn as a nonlinear line named the item characteristic curve, which presents a visual representation of item properties that can be applied in test development and refinement. In addition, the item characteristic curves of several items in a test can be

combined to generate a *test characteristic curve*, which can be used to calculate the performance of examinees at specific ability levels (DeMars, 2010).

IRT Models

In the psychometric measurement field, the word *model* can be described as a "mathematical model in which independent variables are combined numerically to optimally predict a dependent variable" (Embretson & Reise, 2000, p.41). IRT is model-based measurement that specifies how both latent trait and testing items are related to examinee's response (Embretson & Reise, 2000). A large number of models have been developed in the IRT field. Several models are explained as a result of their application in the current study. The one-parameter logistic model (1PL) is the simplest model in IRT developed by Rasch (1960), also known as the Rasch Model. This model is for dichotomous items and functions between latent traits and the item difficulty. The Partial Credit Model (PCM) was developed by Masters (1982), which can be seen as an extensive variety of the Rasch model. This model analyzes test items with multiple steps, and partial credit is assigned for completing several steps in the solving process (Embretson & Reise, 2000). PCM is appropriate for analyzing items with more than two levels of response regarding testing cognition, attitude and personality (Embretson & Reise, 2000; Wu, Adams, Wilson, & Haldane, 2007).

Unidimensionality is the basic assumption for most of the IRT models (Hambleton & Murray, 1983). The application of the unidimensional parametric IRT models needs to assume that the construct being measured is unidimensional (Edelen & Reeve, 2007; Hambleton & Jones, 1993). Some researchers suggest that many tests are designed from several sub-traits which were assumed to evaluate different abilities

(Ansley & Forsyth, 1985). Therefore, multidimensional models were developed to solve this problem. Rasch has developed a multidimensional extension from his basic model (Rasch, 1961), and Kelderman and Rijkes (1994) and Meiser (1996) developed the multidimensional partial credit model. The current study employed a multidimensional Rasch model, Multidimensional Random Coefficient Multinomial Logit Model (Adams, Wilson, & Wang, 1997), called the two-dimensional Rasch Partial Credit Model (2D-RPCM).

Item Fit Statistics

Item fit statistics, including weighted fit and unweighted fit indices, provide information regarding discrepancies in responses (De Ayala, 2009). These item fit indices are used to indicate how well items fit the mathematical model. A misfit item might measure a latent trait that the instrument does not intend to measure. The unweighted fit indicator, also known as outfit, is an outlier-sensitive fit statistic, focusing on an unexpected answer far from a person's ability or an item's difficulty (Linacre, 2015). A very high/low outfit value in an item may mean the item is too difficult/easy for the examinees. Weighted fit indicator, also known as infit, is an inlier-pattern-sensitive fit statistic, which is sensitive to unexpected response patterns to items (Linacre, 2015). A misfit situation in weighted fit may indicate the item measures a different underlying construct from other well fit items.

Item Difficulty

In IRT, item difficulty estimates are indicated in logits. The average of the item difficulty estimates has been traditionally set as a logit value of 0. Person ability (i.e., θ) is estimated relating to item difficulty estimates, so each item and person ability can be

located on a common logit scale (Bond & Fox, 2015). When a person's ability is at the same point of the logit scale as the item difficulty is, the probability of succeeding on the item is 50%. A higher value of item difficulty indicates a decreased likelihood to score on this item. Generally, the more positive logits a person has, the higher ability he/she has; the more positive logits an item is, the more difficult is the item.

Reliability

The person reliability index represents the level of being replicable for person ordering (Wright & Masters, 1982). Person reliability is provided by the Rasch model to inform whether there are enough items spread along the difficulty continuum and enough spread of ability among persons (Bond & Fox, 2015). A sound person reliability needs items that are well aimed on the targeted latent trait, and also a wide range of distribution of ability across the sample that demonstrates a hierarchy of ability on the targeted latent trait (Bond & Fox, 2015).

Item Information Functions

One of the fundamental features of IRT models is the concept of psychometric information (Embretson & Reise, 2000). Higher item information represents a lower standard error of measurement and higher reliability (DeMars, 2010). The amount of psychometric information presenting at all points along the latent ability continuum can be transferred into an *item information curve* (Embretson & Reise, 2000). However, classical test theory assumes that measurement precision is consistent across the continuum of the level of the latent abilities, whereas IRT argues that measurement precision may not be constant (Fraley, Waller, & Brennan, 2000). Item information function is often employed in test development, which allows the test developer to select

items based on the contribution of each test item to the test information function (Hambleton & Jones, 1993).

Items with higher discriminating power may provide more information, and items are most informative when the item difficulty (b) is closed to the latent trait level (θ) (Embretson & Reise, 2000; Fraley et al., 2000; Hambleton & Jones, 1993). The construction of an item information curve is based on the calculation of item information of a particular item on all levels of trait values. By evaluating the item information characteristic curves, items that contribute little information, implying to contribute little to precision, may be adjusted or discarded (Cai, Thissen, & Du Toit, 2011).

Differential Item Functioning

Differential item functioning (DIF) is used as a technique to examine an instrument at the item level to detect whether there is potential bias. DIF measures whether there are differences between two groups with the same level of latent trait (e.g., social-emotional behavior), by modeling the likelihood that items receive scores indicating the target competence (Embretson & Reise, 2000). In a DIF analysis, one of the groups is the focal group, which is the one being investigated, and the other one is the reference group, which serves as the standard to be compared with the focal group. DIF analysis works as Linacre (2015, p.549) noted, "For each slice, a cross-tabulation is constructed for each pair of person classifications against each scored response level. An odds-ratio is computed from the cross-tab."

This chapter reviewed the elemental topics for the current study, defining the topic and providing fundamental knowledge about early childhood social-emotional competence. The assessment section described the measuring approaches for social-

emotional competence. Further, culture was addressed as an important influence on children's development. Lastly, ASQ:SE research and IRT method were reviewed, providing background for the current study in which the psychometric properties of ASQ:SE-2 and ASQ:SE Traditional Chinese version were evaluated using IRT statistics.

CHAPTER III

METHOD OF STUDY

Introduction

The ASQ: SE has been developed as a screening tool used in social-emotional developmental screening for infants and young children. Through the administration of the ASQ:SE, assessment results may assist parents to better understand whether a child's social-emotional status is typical, by comparing the child's score with the cutoff points that were established using a large normative sample. This study investigated the psychometric properties of the English version of the ASQ:SE-2 and the Taiwanese Traditional Chinese translated version (ASQ:SE-TC), using IRT and further comparing the functioning of items in the two versions. The 48-month interval of the ASQ:SE-2, targeting children from 42 months to 54 months old, was used. The research questions include:

1. What are the psychometric properties of ASQ:SE-2 and ASQ:SE-TC using IRT to examine:

    1.1. What is a good-fitting model for the data of the ASQ:SE-2?

    1.2. What are the item fit statistics for the ASQ:SE-2 and ASQ:SE-TC?

    1.3. What are the item difficulty statistics for the ASQ:SE-2 and ASQ:SE-TC?

    1.4. What is the reliability of the ASQ:SE-2 and ASQ:SE-TC?

    1.5. What are the item information curves for the ASQ:SE-2 and ASQ:SE-TC?

2. What is the differential item function (DIF) comparing the ASQ:SE-TC and ASQ:SE-2?

3. What is the difference between the distribution of children's latent traits for the

U.S. and Taiwanese samples?

Participants

This study focused on young children in the U.S. and Taiwan, ages 42 months 0 days to 53 months 30 days. The U.S. sample was retrieved from an ASQ:SE-2 extant database, and Taiwanese sample was recruited for the current study. Approval for research with human research participants was obtained through the university via the institutional review board prior to beginning the study. Once the institutional review board approved the study, recruitment was begun in Taiwan.

United States

For the U.S. sample, the current analysis used the extant dataset which contained 3,005 children/parents or caregivers dyads who participated as a part of a national normative study of the ASQ:SE Second Edition (Squires et al., 2015). The recruitment of families was strategically conducted so that the sample roughly reflected the U.S. 2010 census data on variables including race/ethnicity, mother's education level, and family income. Recruitment occurred through notices on online webpages (i.e., http://asq.uoregon.edu/ and http://asqoregon.com/). The demographic characteristics of this sample are presented in Table 1.

Table 1. Demographic characteristics of the U.S. sample for the ASQ:SE-2 (*N* = 3,005).

Characteristic	*n*	%	Characteristic	*n*	%
Gender			Family Income (USD)		
Boy	1,815	60.4	0-12,000	213	7.1
Girl	1,190	39.6	12,001-24,000	309	10.3
Ethnicity			24,001-40,000	441	14.7
Asian	107	3.6	Over 40,000	1,734	57.7
White	2,001	66.6	Undisclosed	308	10.2
Native American	23	0.8			

Table 1. (continued).

Characteristic	*n*	%	Characteristic	*n*	%
Hawaiian	2	0.1	Author		
Black	153	5.1	Mother	2,222	73.9
Hispanic	157	5.2	Father	167	5.6
Pacific Islander	8	0.3	Guardian	13	0.4
Other	6	0.2	Grand parents	51	1.7
Don't know	2	0.1	Foster parent	34	1.1
Mixed	244	8.1	Both parents	106	3.5
Undisclosed	302	10.0	Other	62	2.1
Mother's education			Teacher	149	5.0
Less than high school	105	3.5	Adoptive parent	51	1.7
High school	744	24.8	Childcare provider	113	3.8
AA degree	436	14.5	Undisclosed	37	1.2
College or higher	1,533	51.0			
Don't know	153	5.1			
Undisclosed	34	1.1			

Taiwan

For the Taiwanese sample, 500-2,000 child/parent dyads were targeted, to obtain accurate parameter estimates using multidimensional IRT (Ackerman, 1994; Jiang, Wang, & Weiss, 2016; Kose & Demirtasli, 2012). Variables including regions, children's gender, parents' education, and income were used as guidelines so that the sample would reflect the Taiwanese 2015 census (Ministry of Health and Welfare, 2016; Ministry of the Interior, 2016). Other demographic variables were asked including rater's relationship with the child, mother's age, the common language in the family, and the immigrant family identity. Parents were recruited using paper-pencil and online versions of the 48-month ASQ:SE-2 Traditional Chinese.

Paper-pencil copies were collected by the following procedure. First, public and private kindergarten principals were contacted to ask if their classrooms were willing to

assist in this study. Once they consented to participate, the corresponding number of questionnaires was sent to the kindergartens. The principals received training on the phone or face-to-face by a professional who had master's degree in early intervention and was familiar with the questionnaire. The principal then trained teachers to use the questionnaires, so that they could instruct parents how to complete the ASQ:SE-2 Traditional Chinese version. Teachers then delivered the questionnaires to parents. Teachers took the responsibilities to describe how to administer the questionnaires to parents and collect the questionnaires. Each teacher was rewarded a gift card of $200 in New Taiwan Dollar (NTD) and each parent was rewarded a gift card of $50 NTD.

An online version of the ASQ:SE-TC was set up on a Qualtrics website. The link to the questionnaire was posted on parenting websites and Facebook. A Facebook advertisement was bought for recruitment, first targeting each county/city in Taiwan. The targeted county/city then was changed based on the portion of the population in the five regions that had been collected, attempting to match the Taiwanese 2015 census. Each parent who completed the online copy was asked if he/she was interested in a gift card of $50 NTD, and a gift card was then delivered by mail.

Measures

Ages & Stages Questionnaires: Social-Emotional Second Edition (ASQ:SE-2)

The 48-month interval of ASQ:SE-2 was the primary outcome measure. This interval has 36 items and each item has three response options that allow parents or primary caregivers to observe their children and indicate the frequency of their children's social-emotional skills or concerns (i.e., "Often or Always," "Sometimes," or "Rarely or Never"). Answers receive numeric values reflecting competence (0 points) or problem

behaviors (5 or 10 points). This study included only the first 35 items because Item 36 is an open-ended question (i.e., *"Has anyone shared concerns about your child's behaviors? If 'sometimes' or 'often or always', please explain:"*). Furthermore, an additional response option with each item allowed parents to indicate if the behavior is a concern; however, the "concern" option (with an associated value of 5 points) was not included in this analysis because this response is not part of the scoring options that evaluate the frequency of behaviors.

Ages & Stages Questionnaires: Social-Emotional Traditional Chinese (ASQ:SE-TC)

ASQ:SE-2 was translated into Traditional Chinese by this researcher and a professional who are both proficient in both English and Chinese, with careful consideration to cultural appropriateness and linguistic meaning. Items of the ASQ:SE-2 appeared appropriate for the Taiwanese culture after our review. The only item that needed cultural adaption was, "*Does your baby like to play games like peek-a-boo?*" appearing on the 12-month interval. In Taiwan, a parenting game similar to peek-a-boo is called "duǒ māo māo" which means "hiding kitty kitty;" this game was substituted instead.

Items translated into Traditional Chinese were then back translated to English by an independent translation company based in Taiwan to ensure that the translation did not differ from the original items. Any differences between the original ASQ:SE and the back translation manuscript were compared. The identified differenced in items were discussed and adjusted. The results indicated the translation well reflected the original English version.

Data Analysis

The data analysis procedure consisted of four phases. The beginning phase was to evaluate which IRT model fit ASQ:SE-2 the U.S. sample. Once the model with better fit was identified, it was used to analyze the Taiwanese sample. Secondly, psychometric

properties of ASQ:SE-2 and ASQ:SE-TC were examined including item fit, item difficulty, reliability, and item information function. Third, items of ASQ:SE-2 and ASQ:SE-TC were compared through DIF analyses using the model identified in the first phase. Fourth, consistency in distributions of the U.S. and Taiwanese children's social-emotional trait were examined. Analyses were conducted using ConQuest 4.5 (Adams, Wu, & Wilson, 2016).

Model Evaluation

Model evaluation for the ASQ:SE-2 included identifying which IRT model would accurately reflect the relations of children's social-emotional competence and the characteristics of items. The assumption of unidimensionality ensures that test items are targeting the same latent trait. If a test consists of more than one dimension, multidimensional IRT models can be used. Based on the factorial structure study in the ASQ:SE first edition (Chen, Filgueiras, Squires, & Landeira-Fernandez, 2016), a single-construct structure was proposed to compare with a two-construct structure in each age interval. The two-construct structure was established according to the theory that social competence and emotional competence are considered as two distinct but highly related areas (Campos et al., 1994; Raver & Zigler, 1997; Squires et al., 2015). The results suggested the two-construct structure, labeled Sociality and Emotion, presented better fit indices than the single-construct structure did.

Since three new items were added to the 48-month interval of ASQ:SE-2, the dimensional structure had to be re-evaluated. The current study categorized the three new items, Item 33-35, into the previous two-construct structure (see Table 2), and compared this two-construct structure to the one-construct structure using IRT models. The single-

construct structure was analyzed using the one-dimensional Rasch Partial Credit Model (1D-RPCM), while the two-construct structure was analyzed using two-dimensional Rasch Partial Credit Model (2D-RPCM) (Adams et al., 1997).

Table 2. Items of the 48-month interval of the ASQ:SE-2 and the dimensions for 2D-RPCM.

Item	Item description	Dimension
2	Does your child cling to you more than you expect?	1
4	When upset, can your child calm down within 15 minutes?	1
6	Does your child seem too friendly with strangers?	1
7	Can your child settle himself down after periods of exciting activity?	1
8	Does your child cry, scream, or have tantrums for long periods of time?	1
11	Does your child have eating problems?	1
13	Does your child do what you ask her to do?	1
14	Does your child seem happy?	1
15	Does your child sleep at least 8 hours in a 24-hour period?	1
16	Does your child seem more active than other children his age?	1
18	Can your child stay with activities he enjoys for at least 10 minutes?	1
20	Can your child move from one activity to the next with little difficulty?	1
22	Does your child do things over and over and can’t seem to stop?	1
23	Does your child hurt himself on purpose?	1
24	Does your child follow rules?	1
25	Does your child destroy or damage things on purpose?	1
31	Does your child try to hurt other children, adults, or animals?	1
32	Does your child show an unusual interest of sexual language and activity?	1
33	Does your child wake three or more times during the night?	1
34	Is your child too worried or fearful?	1
1	Does your child look at you when you talk to him?	2
3	Does your child talk and/or play with adults she knows well?	2
5	Does your child like to be hugged or cuddled?	2
9	Is your child interested in things around her, such as people, toys, and foods?	2
10	Does your child stay dry during the day?	2

Table 2. (continued).

Item	Item description	Dimension
12	Do you and your child enjoy mealtimes together?	2
17	Does your child use words to tell you what she wants or needs?	2
19	Does your child use words to describe her feelings and the feelings of others?	2
21	Does your child explore new places, such as a park or a friend's home?	2
26	Does your child stay away from dangerous things?	2
27	Can your child name a friend?	2
28	Does your child show concern for other people's feelings?	2
29	Do other children like to play with your child?	2
30	Does your child like to play with other children?	2
35	Does your child have simple back-and-forth conversations with you?	2

Note. Dimension 1 = Emotion; Dimension 2 = Sociality.

The model fit comparison for the unidimensional and multidimensional models was made based on deviance value. The comparison of the deviance value of the two nested models provided information about whether the 2D-RPCM should be rejected. Deviance is a measure that indicates how well the item response model has fit the data (Wu et al., 2007). Comparing the deviance difference (i.e., 1D-RPCM minus 2D-RPCM) between nested models to a chi-squared distribution with corresponding degrees of freedom, if the deviance of 1D-RPCM is greater and significant, it can be concluded that the fit of 2D-RPCM is significantly better than the fit of 1D-RPCM.

Psychometric Properties

The psychometric properties of ASQ:SE-2 and ASQ:SE-TC were examined using the better fit model identified through the previous evaluation and were presented corresponding to the research questions (i.e., 1.2 to 1.5) regarding item fit statistics, item difficulty, reliability, and item information curves.

Item fit statistics. Weighted fit and unweighted fit statistics were presented as mean square standardized residuals (MNSQ). The residuals indicate the difference between the value predicted by the model and the observed value calculated from empirical data. With higher squared residual, a larger misfit exists between the model and the data. A misfit situation in weighted fit may indicate the item measures a different underlying construct from other well fit items. The weighted fit statistic is more often used to evaluate the quality of items than unweighted fit for users of the Rasch model (Bond & Fox, 2015), and it is suggested that values ranging from 3/4 to 4/3 of the ideal value, 1.00 for the Rasch model, is an acceptable MNSQ range of 0.75-1.33 (Wu, Adams, & Wilson, 1998).

Item difficulty. In IRT, a higher value of item difficulty indicates a decreased likelihood to score on this item. In the ASQ:SE-2 and ASQ:SE-TC, a higher value of item difficulty indicates a decreased likelihood to score on an item regarding a problem behavior, for which a low score is desirable. (That is, on the ASQ:SE, higher scores indicate more behavior problems.) According to the total score distribution in the data set from the original normative study in the U.S., it was positively skewed, with the majority of children receiving low scores (Squires et al., 2015). Through IRT analyses, the same trend was expected for children who were assessed using ASQ:SE-2 and ASQ:SE-TC, which meant items were expected to be difficult for children (i.e., expecting a high and positive logit value).

Reliability. The precision of person estimates can be examined by the "expected a posteriori/plausible value (EAP/PV)" reliability provided by the ConQuest software. EAP/PV reliability, the ratio of the variance of the expected a posteriori trait values to

total person variance of the latent traits, or can be simply explained as the ratio of modeled variance to observed variance (Scalise, 2012). The evaluation of EAP/PV reliability is similar to Cronbach's alpha. By examining EAP/PV reliability, the replicability of person placement across items measuring the same content can be claimed (Bond & Fox, 2015).

Item information curves. The information function curves indicate how well each latent trait level is being estimated, which are traditionally shown in a graphic figure. Items with higher information represent lower standard error of measurement and higher reliability (DeMars, 2010). The x-axis represents the ability which is being be measured, and the y-axis represents the amount of information. Each of the curves represents an item. A figure regarding item information curves describes that when the latent trait is in a certain level, a specific item maybe possible to reveal more/less information than other items. Explanation of the information curves relied on visual analyses.

Differential Item Functioning

DIF technique was used to examine whether ASQ:SE items functioned differently for the U.S. sample and Taiwanese sample. DIF measures whether the two groups of examinees with the same level of latent trait reply to an item differently. In the current study, the U.S. sample was the reference group and the Taiwanese sample the focal group. Two datasets were compared with each other and the magnitude of DIF was evaluated to determine if the effect of DIF was substantive, by applying the Educational Testing Service (ETS) standard that $|DIF| \leq 0.43$ logits (level A) is considered negligible, 0.64 logits $\geq |DIF| \geq 0.43$ logits (level B) is slight to moderate, and $|DIF| \geq 0.64$ logits is moderate to large (Zwick, Thayer, & Lewis, 1999).

Comparison of Theta

In IRT analysis, each examinee's ability (i.e., theta) was estimated in addition to item difficulty. A comparison of the distribution of each person's social-emotional trait for the U.S. and Taiwanese sample was made. First, two datasets were combined into one and then calibrated using the IRT model. Secondly, the person distributions of theta estimates on the latent traits were presented as histograms by country. Visual comparison was made to inspect whether the distributions in two cultural groups were consistent.

This chapter explained the methodology that was applied in the current study, including the approaches to evaluate models, the indices to examine psychometric properties, the DIF technique to inspect the different probability to perform the items between two groups. Lastly, the method used to compare the theta distributions of the U.S. sample and Taiwanese sample was explained.

CHAPTER IV

RESULTS

Participants

This study included two samples from different countries, the U.S. and Taiwan. The U.S. sample was retrieved from an extant dataset collected between March 2010 and October 2015 as part of a national normative study of the ASQ:SE-2, with demographic information presented in Table 1. The Taiwanese sample was recruited between August and December 2016 using paper-pencil and online versions of the 48-month ASQ:SE-2 Traditional Chinese.

Demographic Information of Taiwanese Sample

The survey of demographic information asked about geographic regions, children's gender, parents' education, low income qualification, rater's relationship with the child, mother's age, the common language in the family, and the immigrant family identity.

Paper-pencil and online versions of the 48-month ASQ:TC were completed by participants. A total of 1,455 parents/caregivers completed questionnaires, including paper-pencil (n = 444) and online (n = 1,011) versions. A total of 500 hard copies were sent to the kindergartens and 444 copies were returned, with a response rate of 88.8%. Links to the online version were disseminated via posting information on parenting websites, parenting groups on Line software, as well as Facebook (e.g., 28,368 users were reached by the Facebook advertisement). A total of 1,786 responses were collected online. Eight respondents dropped out after agreeing to participate; a number of 350 respondents replied that their children were not in the target age range (i.e., 42 months 0

days to 53 months 30 days); the data completed by 417 respondents were excluded since they left more than one-third of the items empty (i.e., more than 12 items). Thus, 1,011 out of 1,786 responses (56.6%) remained in the current analysis. The demographic characteristics of this sample are presented in Table 3.

Table 3. Demographic characteristics of the Taiwan sample for the ASQ:SE-TC (*N* = 1,455).

Characteristic	Paper-pencil (*n* = 444) *n* (%)	Online (*n* = 1,011) *n* (%)	Combined (*N* = 1,455) *N* (%)	2015 Census %	[a]Difference %
Sex					
Boy	233 (52.5)	557 (55.1)	790 (54.3)	51.9	+2.4
Girl	206 (46.4)	454 (44.9)	660 (45.4)	48.1	-2.7
Undisclosed	5 (1.1)	0 (0)	5 (0.3)		
Mother's education					
Elementary school	3 (0.7)	1 (0.1)	4 (0.3)	1.5	-1.2
Middle school	15 (3.4)	5 (0.5)	20 (1.4)	6.3	-4.9
High school	150 (33.8)	99 (9.8)	249 (17.1)	27.0	-9.9
College or higher	271 (61.0)	901 (89.1)	1,172 (80.5)	65.2	+15.3
Don't know	3 (0.7)	5 (0.5)	8 (0.5)		
Undisclosed	2 (0.5)	0 (0)	2 (0.1)		
Father's education					
Elementary school	1 (0.2)	0 (0)	1 (0.1)	0.3	-0.2
Middle school	22 (5.0)	14 (1.4)	36 (2.5)	7.3	-4.8
High school	137 (30.9)	147 (14.5)	284 (19.5)	31.2	-11.7
College or higher	263 (59.2)	841 (83.2)	1,104 (75.9)	61.3	+14.6
Don't know	4 (0.9)	9 (0.9)	13 (0.9)		
Undisclosed	17 (3.8)	0 (0)	17 (1.2)		
Region					
Northern TW	118 (26.6)	550 (54.4)	668 (45.9)	48.3	-2.4
Central TW	167 (37.6)	238 (23.5)	405 (27.8)	25.1	+2.7
Southern TW	137 (30.9)	194 (19.2)	331 (22.7)	23.5	-0.8
Eastern TW	15 (3.4)	11 (1.1)	26 (1.8)	2.1	-0.3
Outlying islands	7 (1.6)	10 (1.0)	17 (1.2)	1.0	+0.2
Others		8 (0.8)	8 (0.5)		
Family income					
Normal	401 (90.3)	966 (95.5)	1,367 (94.0)	96.9	-2.9
Low income	23 (5.2)	31 (3.1)	54 (3.7)	3.1	+0.6
Don't know	15 (3.4)	14 (1.4)	29 (2.0)		
Undisclosed	5 (1.1)	0 (0)	5 (0.3)		

Table 3. (continued).

Characteristic	Paper-pencil (n = 444) n (%)	Online (n = 1,011) n (%)	Combined (N = 1,455) N (%)	2015 census %	[a]Difference %
Author					
Mother	336 (75.7)	852 (84.3)	1,188 (81.6)		
Father	47 (10.6)	68 (6.7)	115 (7.9)		
Grand parents	15 (3.4)	5 (0.5)	20 (1.4)		
Other relatives	1 (0.2)	24 (2.4)	25 (1.7)		
Teacher	36 (8.1)	56 (5.5)	92 (6.3)		
Others	1 (0.2)	6 (0.6)	7 (0.5)		
Undisclosed	8 (1.8)	0 (0)	8 (0.5)		
Language					
Mandarin	412 (92.8)	1,006 (99.5)	1,418 (97.5)		
Taiwanese	233 (52.5)	317 (31.4)	550 (37.8)		
Hakka	7 (1.6)	14 (1.4)	21 (1.4)		
Formosan languages	3 (0.7)	2 (0.2)	5 (0.3)		
English	3 (0.7)	38 (3.8)	41 (2.8)		
German	0 (0)	1 (0.1)	1 (0.1)		
Vietnamese	1 (0.2)	1 (0.1)	2 (0.1)		
Japanese	1 (0.2)	4 (0.4)	5 (0.3)		
Korea	0 (0)	1 (0.1)	1 (0.1)		
French	1 (0.2)	0 (0)	1 (0.1)		
Shanghai dialect	0 (0)	1 (0.1)	1 (0.1)		
Mother’s age					
20-24	4 (0.9)	6 (0.6)	10 (0.7)		
25-29	37 (8.3)	63 (6.2)	100 (6.9)		
30-34	124 (27.9)	376 (37.2)	500 (34.4)		
35-39	179 (40.3)	459 (45.4)	638 (43.8)		
40-44	67 (15.1)	86 (8.5)	153 (10.5)		
45-49	10 (2.3)	9 (0.9)	19 (1.3)		
50-54	1 (0.2)	0 (0)	1 (0.1)		
Undisclosed	22 (5.0)	12 (1.2)	34 (2.3)		
Immigrant family					
No	416 (93.7)	986 (97.5)	1,402 (96.4)		
Yes	16 (3.6)	24 (2.4)	40 (2.7)		
Don’t know	5 (1.1)	1 (0.1)	6 (0.4)		
Undisclosed	7 (1.6)	0 (0)	7 (0.5)		

Note: [a]Subtracted the percentage in 2015 census column from the percentage in Combined column.

Based on demographic information, 78.2% mother’s age were 30 to 39 years old, thus Taiwanese 2015 census for this age group was used (see Table 3). Five regions as defined by the Taiwanese government were used as targets, including Northern Taiwan

(including Taipei, New Taipei, Taoyuan, Yilan, Hsinchu City, Hsinchu County, and Keelung), Central Taiwan (including Taichung, Miaoli, Changhua, Yunlin, and Nantou), Southern Taiwan (including Kaohsiung, Tainan, Pingtung, Chiayi County, and Chiayi City), Eastern Taiwan (including Hualien and Taitung), and Outlying Islands (including Kinmen, Lianjiang, and Penghu). The regional ratio of population listed on Table 3 were based on the 0-6 years old children population in Taiwanese 2015 census (Ministry of the Interior, 2016).

The demographic survey in this study asked parents if their family qualified as a low or medium-low income family, as it is a sensitive question to ask about the family's income in Taiwan. This sensitive question might cause parents to be offended or displeased. Efforts were made to mirror national demographic information regarding poverty (Ministry of Health and Welfare, 2016). A family was labeled as a low income family when the total income of the family divided by the number of family members was lower than $10,869 NTD per month; and a family qualified as the medium-low income family when the total income of the family divided by the number of family members was lower than $16,304 NTD per month (My E Government, 2015).

Author is a variable used to ask who completed the questionnaire. The common languages used in the family were also asked, including Mandarin (i.e., the official language of Taiwan), Taiwanese, Hakka, Formosan languages (i.e., the languages of the indigenous peoples of Taiwan), and others. A family could check more than one language.

The difference between the combined dataset (i.e., adding paper-pencil and online data) and the 2015 census indicated that the combined dataset well represented the demographic variables of the Taiwan 2015 census (see Table 3) including gender (i.e.,

boy = +2.4%; girl = -2.7%), regions (i.e., -2.4% to +2.7%), and low and low-medium income family (i.e., +0.6%). However, the combined dataset overrepresented the education level of the mother (i.e., +15.3%) and father (i.e., +14.6%) with college or higher education level.

Model Evaluation

The model evaluation for the unidimensional and multidimensional models was made based on comparing the deviance values. Deviance is a statistic indicating how well the model fits the data. The difference between deviances of models (i.e., 1D-RPCM minus 2D-RPCM), referring to a chi-squared distribution with corresponding degrees of freedom, could determine which model is the better fit.

The deviance in 1D-RPCM and 2D-RPCM for the 48-month interval of the ASQ:SE-2 (i.e., using U.S. sample) was estimated as shown in Table 4. The deviance of parameter equality of 1D-RPCM was 121,543.61 and the total number estimated parameters was 71; the chi-square test of parameter equality of 2D-RPCM was 121,085.54 and the total number estimated parameters was 73, adding mean and variance of the extra dimension. The difference between the deviance of 1D-RPCM and 2D-RPCM was distributed as a chi-square with two degrees of freedom. The estimated deviance difference between the models was 458.07. This deviance value lay beyond 13.82 ($p < .001$), since a p-value of less than .001 was greater than the conventionally accepted significance level of 0.05 ($\chi^2 = 5.991$). There was a statistically significant difference in applying 2D-RPCM as used here for this dataset as compared to the 1D-RPCM.

The model evaluation indicated that 2D-RPCM fit the U.S. data better than 1D-RPCM. The comparison was then made to compare the models using Taiwanese subjects. The result is also shown in Table 4, consistent with results for the U.S. sample.

Table 4. Comparison of model fit statistics between 1D-RPCM and 2D-RPCM by country.

Country	*n*	1D-RPCM		2D-RPCM		Deviance Difference	p^a
		Deviance	Parameters	Deviance	Parameters		
U.S.	3,005	121,543.61	71	121,085.54	73	458.07	$p < .001$
Taiwan	1,455	60,021.20	71	59,815.59	73	205.60	$p < .001$

Note. $^a\chi^2 = 13.82$

Psychometric Properties

The psychometric properties of the ASQ:SE-2 and ASQ:SE-TC, including item fit statistics, item difficulty, reliability, and item information functions, were investigated by 2D-RPCM, the better fit model based on the previous evaluation.

Descriptive Statistics

IRT analysis focuses on categorical or ordinary data. The information regarding frequency and percentage of each item's options (i.e., 0, 5, 10 points) are listed in Table 5.

Table 5. Frequency and percentage of the categories for ASQ:SE-2 and ASQ:SE-TC items.

Item	ASQ:SE-2 (*N* = 3,005)						ASQ:SE-TC (*N* = 1,455)					
	0		5		10		0		5		10	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
1	2,106	70.4	791	26.4	96	3.2	1,089	74.8	350	24.1	16	1.1
2	1,591	53.2	1,089	36.4	310	10.4	142	9.8	782	53.9	528	36.4
3	2,626	87.5	315	10.5	61	2.0	1,227	84.3	202	13.9	26	1.8
4	2,187	73.0	690	23.0	118	3.9	964	66.3	445	30.6	44	3.0
5	2,338	78.0	589	19.6	72	2.4	1,208	83.1	219	15.1	26	1.8
6	1,621	54.1	966	32.2	412	13.7	800	55.1	499	34.3	154	10.6
7	1,524	51.0	1,245	41.6	222	7.4	727	50.1	649	44.8	74	5.1
8	1,673	55.9	980	32.7	340	11.4	771	53.2	585	40.4	92	6.4

Table 5. (continued).

Item	ASQ:SE-2 (*N* = 3,005)						ASQ:SE-TC (*N* = 1,455)					
	0		5		10		0		5		10	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
9	2,679	89.6	278	9.3	32	1.1	1,256	86.6	182	12.5	13	0.9
10	2,465	82.6	310	10.4	211	7.1	1,180	82.5	136	9.5	115	8.0
11	2,457	82.2	315	10.5	218	7.3	1,168	81.9	185	13.0	74	5.2
12	2,156	72.0	717	23.9	123	4.1	748	52.1	633	44.1	55	3.8
13	1,377	45.9	1,472	49.1	149	5.0	955	66.5	460	32.0	21	1.5
14	2,607	87.2	371	12.4	13	0.4	1,269	88.6	158	11.0	6	0.4
15	2,741	92.0	195	6.5	44	1.5	1,347	93.7	82	5.7	8	0.6
16	1,210	40.4	1,089	36.3	697	23.3	325	22.6	670	46.7	441	30.7
17	2,344	78.4	552	18.5	94	3.1	1,276	89.1	139	9.7	17	1.2
18	2,209	74.0	630	21.1	148	5.0	1,206	84.0	201	14.0	29	2.0
19	2,202	73.4	525	17.5	274	9.1	1,207	84.0	187	13.0	43	3.0
20	1,757	58.7	1,018	34.0	218	7.3	924	64.5	481	33.6	27	1.9
21	2,323	77.8	572	19.2	89	3.0	979	68.3	404	28.2	50	3.5
22	2,356	78.5	428	14.3	216	7.2	982	69.0	333	23.4	109	7.7
23	2,704	90.3	234	7.8	55	1.8	1,331	92.9	79	5.5	23	1.6
24	1,642	54.8	1,204	40.2	149	5.0	858	60.0	536	37.5	37	2.6
25	2,079	69.4	737	24.6	181	6.0	919	63.8	475	33.0	46	3.2
26	2,410	80.3	438	14.6	152	5.1	1,129	78.1	239	16.5	77	5.3
27	2,450	82.0	298	10.0	241	8.1	1,266	87.6	128	8.9	51	3.5
28	2,092	70.0	709	23.7	187	6.3	882	61.0	482	33.4	81	5.6
29	2,187	72.9	688	22.9	125	4.2	1,062	73.4	365	25.2	19	1.3
30	2,310	77.0	579	19.3	110	3.7	1,150	80.5	253	17.7	26	1.8
31	2,158	72.0	686	22.9	155	5.2	1,097	75.8	307	21.2	43	3.0
32	2,899	96.9	62	2.1	30	1.0	1,256	87.0	165	11.4	23	1.6
33	2,476	83.4	392	13.2	101	3.4	1,220	84.8	187	13.0	32	2.2
34	1,846	81.1	340	14.9	90	4.0	1,148	79.7	234	16.3	58	4.0
35	1,529	79.4	240	12.5	157	8.2	1,246	86.2	143	9.9	57	3.9

Item Fit Statistics

Item weighted fit was applied to examine how well the 2D-RPCM as well as the items fit the current datasets. The ideal value of MNSQ for the Rasch model is 1.00, and the acceptable MNSQ range is 0.75-1.33 (Wu, Adams, & Wilson, 1998). A fit value higher than 1.33 is more problematic than one lower than 0.75, as it is a result of unexpected data patterns indicating that the item might not measure the latent trait reliably.

The results indicated that 3% (1 of 35, Item 6) did not fit 2D-RPCM for the U.S. dataset, while all items fit 2D-RPCM for the Taiwanese dataset. The weighted fit statistics for each item can be seen in Table 6, and other statistics including the range, mean, and standard deviation of weighted fit can be found in Table 7.

Table 6. Item difficulty and item fit statistics for ASQ:SE-2 and ASQ:SE-TC items.

Item	ASQ:SE-2 (*N* = 3,005)				ASQ:SE-TC (*N* = 1,455)			
	Difficulty	Error	Weighted fit	Unweighted fit	Difficulty	Error	Weighted fit	Unweighted fit
1	2.45	0.07	0.95	0.89	2.82	0.14	0.98	0.97
2	1.19	0.04	1.27	1.35	-0.79	0.06	1.19	1.22
3	3.08	0.08	0.95	1.01	2.73	0.11	1.01	1.02
4	2.07	0.06	0.87	0.79	1.88	0.08	0.94	0.90
5	2.77	0.07	1.14	1.48	2.70	0.11	1.06	1.14
6	1.03	0.04	1.43	1.61	1.05	0.05	1.23	1.34
7	1.34	0.05	0.87	0.85	1.38	0.07	0.95	0.93
8	1.17	0.04	0.91	0.90	1.31	0.06	0.97	0.96
9	3.49	0.10	0.86	0.58	3.14	0.15	0.95	0.80
10	2.22	0.06	1.31	1.79	1.82	0.07	1.26	2.02
11	1.87	0.05	1.10	1.26	1.79	0.07	1.02	1.03
12	2.34	0.06	1.18	1.24	1.75	0.08	1.16	1.18
13	1.48	0.05	0.84	0.83	2.26	0.12	0.92	0.89
14	3.48	0.14	0.90	0.65	3.19	0.21	0.95	0.79
15	2.93	0.09	0.96	0.74	3.11	0.18	0.98	0.92
16	0.42	0.04	1.22	1.28	-0.19	0.05	1.16	1.18
17	2.63	0.07	0.85	0.71	3.04	0.14	0.89	0.56
18	1.96	0.05	0.93	0.84	2.32	0.10	0.95	0.82
19	1.87	0.05	0.92	0.83	2.43	0.09	0.89	0.81
20	1.49	0.05	0.85	0.80	2.11	0.10	0.92	0.88
21	2.64	0.07	1.17	1.38	2.08	0.09	1.08	1.16
22	1.82	0.05	0.98	0.98	1.41	0.06	1.02	1.01
23	2.78	0.08	0.96	0.80	2.55	0.11	0.99	1.00
24	1.64	0.05	0.82	0.78	1.88	0.09	0.89	0.87
25	1.77	0.05	0.94	0.90	1.82	0.08	0.93	0.91
26	2.38	0.06	1.19	1.50	2.00	0.08	1.13	1.46
27	2.12	0.05	0.95	0.81	2.40	0.09	0.95	0.74
28	2.04	0.06	0.93	0.90	1.68	0.07	0.96	0.96
29	2.35	0.06	0.85	0.71	2.70	0.13	0.88	0.80
30	2.51	0.06	0.93	0.88	2.63	0.11	0.95	0.87
31	1.91	0.05	0.93	0.88	2.02	0.09	0.96	0.93
32	3.23	0.10	1.08	1.20	2.48	0.11	1.01	0.97

Table 6. (continued).

Item	ASQ:SE-2 (*N* = 3,005)				ASQ:SE-TC (*N* = 1,455)			
	Difficulty	Error	Weighted fit	Unweighted fit	Difficulty	Error	Weighted fit	Unweighted fit
33	2.32	0.06	1.09	1.36	2.28	0.10	1.02	1.16
34	2.20	0.06	1.13	1.25	1.90	0.08	1.07	1.21
35	2.03	0.06	0.94	0.85	2.31	0.09	0.93	0.98

Item Difficulty

For the ASQ:SE-2 and ASQ:SE-TC, high item difficulty meant that the problem behavior rarely occurred in children, or that social-emotional competence was achieved by most children. The item difficulty estimates for ASQ:SE-2 and ASQ:SE-TC can be found in Table 6, and the range, mean, and standard deviation of item difficulty estimates can be found in Table 7.

Table 7. Descriptive statistics of item difficulty and item fit statistics for the ASQ:SE-2 and ASQ:SE-TC.

Measure	Difficulty				Weighted fit				Unweighted fit			
	Min	*Max*	*M*	*SD*	*Min*	*Max*	*M*	*SD*	*Min*	*Max*	*M*	*SD*
ASQ:SE-2	0.42	3.49	2.14	0.69	0.82	1.43	1.01	0.15	0.58	1.79	1.02	0.30
ASQ:SE-TC	-0.79	3.19	2.06	0.84	0.88	1.26	1.00	0.10	0.56	2.02	1.01	0.25

Furthermore, ConQuest software generated “Wright Map”, showing a graphical representation of children’s social-emotional trait distribution on each dimension. For the ASQ:SE-2 using the U.S sample, the Wright Map (see Figure 1) presented a distribution of the Emotion trait (Dimension 1) in the left panel. The middle panel showed the same for the Sociality trait (Dimension 2). The right panel describes the ordering of item difficulty. Generally, examinees have a lower probability to receive high scores on a difficult item. A higher score gain in ASQ:SE indicates lower social-emotional competence, which is not desirable. A difficult item on ASQ:SE-2 may measure a

relatively less common social-emotional behavior. From the Wright Map, the relations between item difficulty estimates and person ability estimates can be explicit. The Wright MAP for the ASQ:SE-TC is presented in Figure 2.

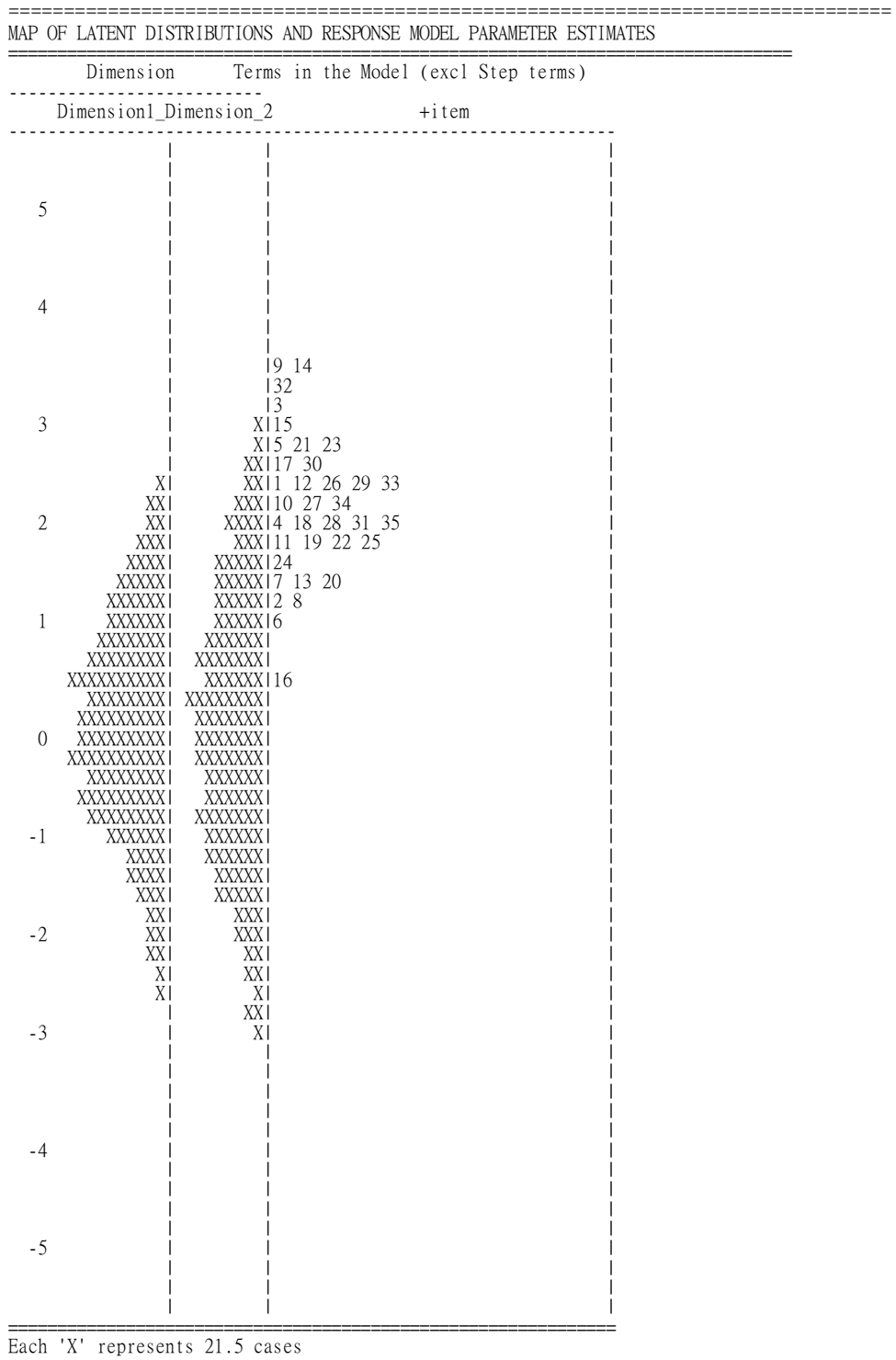

Figure 1. Wright Map for the U.S. sample on the 48-month interval.

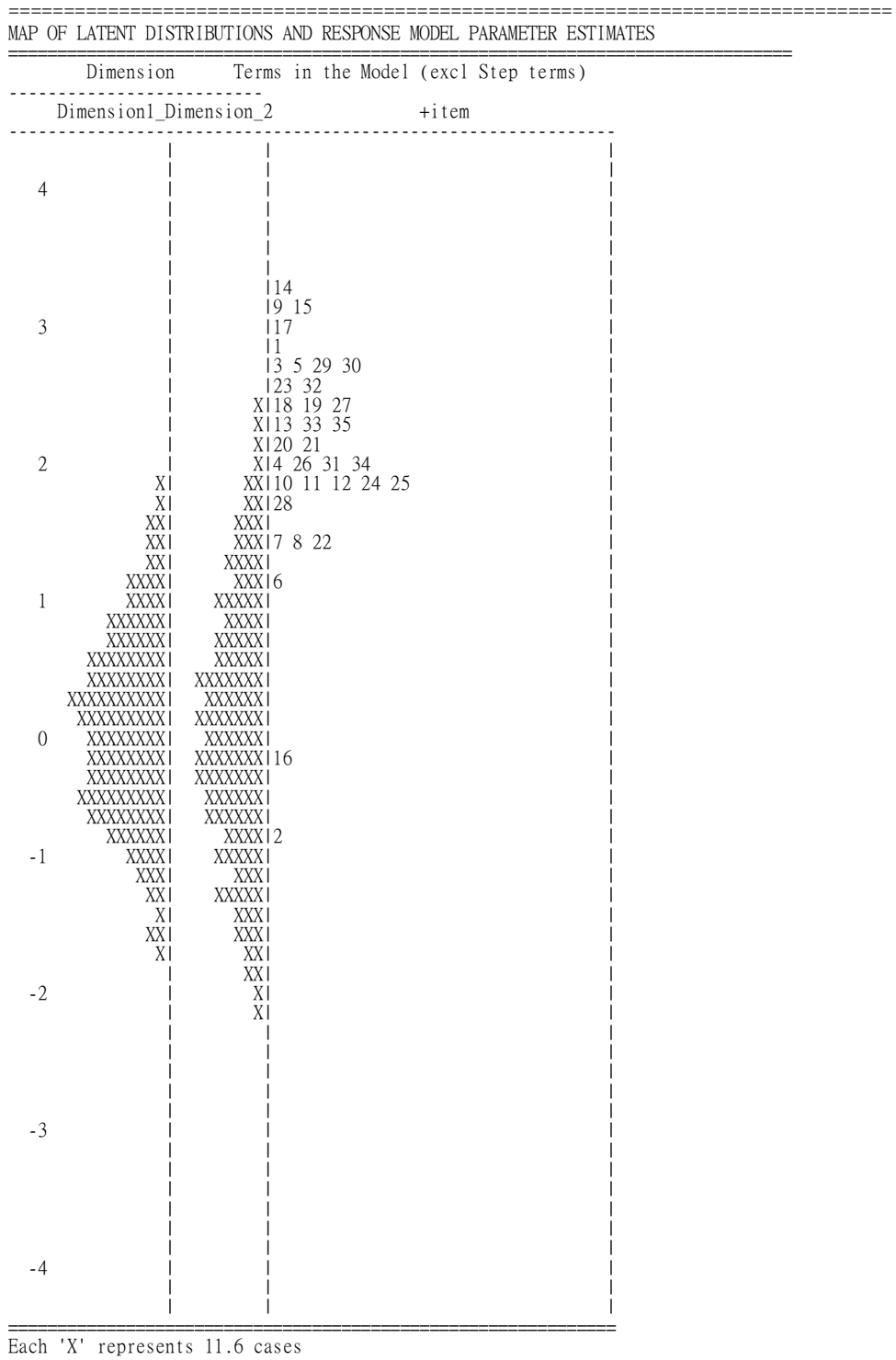

Figure 2. Wright Map for the Taiwanese sample on the 48-month interval.

Reliability

EAP/PV reliability, calculated by ConQuest software, is the ratio of modeled variance to observed variance. The EAP/PV for the ASQ:SE-2 and ASQ:SE-TC can be found in Table 8.

Table 8. EAP/PV reliability for the ASQ:SE-2 and ASQ:SE-TC.

Measure	1D-RPCM	2D-RPCM	
		Emotion (Dimension 1)	Sociality (Dimension 2)
ASQ:SE-2	0.86	0.84	0.81
ASQ:SE-TC	0.79	0.75	0.74

Item Information Functions

Item information curves are presented separately by dimensions. For the U.S sample, items in the Emotion dimension are shown in Figure 3, while items in the Sociality dimension are shown in Figure 4. Item information curves represent the amount of information in each item crossing different levels of social/emotional competence. For instance, in Figure 3, Item 11 “Does your child have eating problems?” informs in the low level of Emotion competence ($\theta = 2.0$) in general, whereas Item 16, “Does your child seem more active than other children his age?” is the most informative item at the moderate level of Emotional competence ($\theta = 0$). This implies that when the latent trait is in a certain level, a specific item maybe possibly reveal more information than other items. Items with higher information represent lower standard error of measurement and the higher reliability (DeMars, 2010). For the Taiwanese sample, Items in the Emotion dimension are shown in Figure 5, while items in the Sociality dimension are shown in Figure 6.

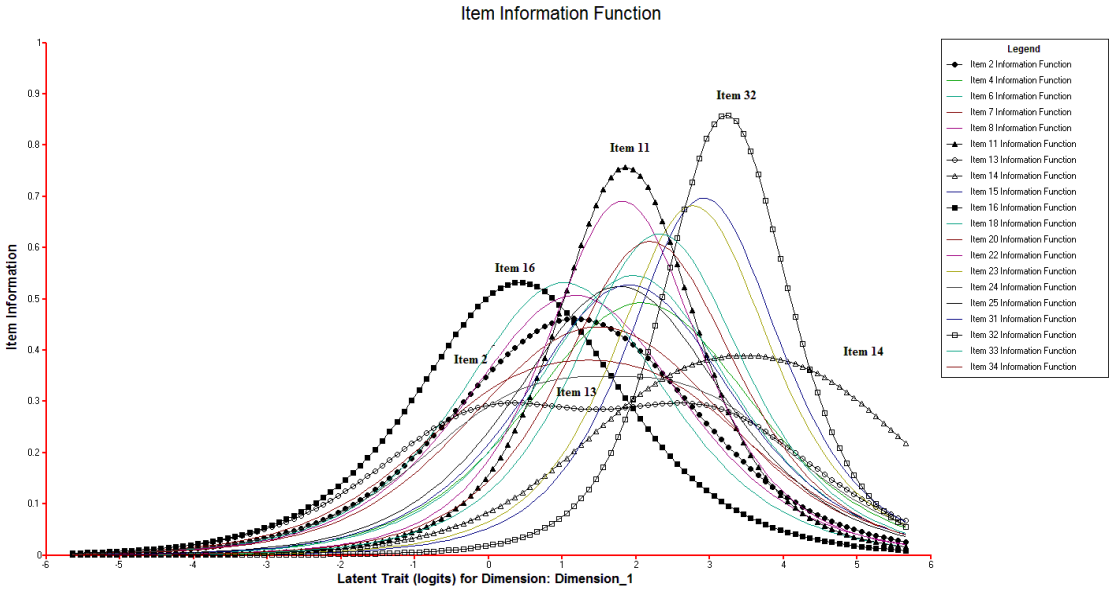

Figure 3. Item information function for Emotion trait (U.S.).

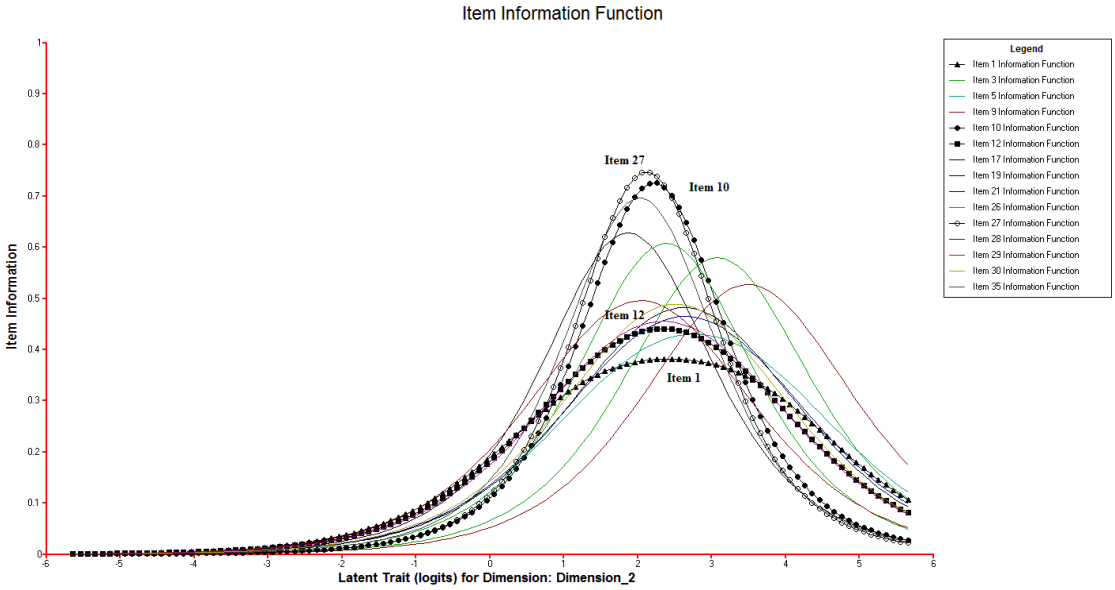

Figure 4. Item information function for Sociality trait (U.S.).

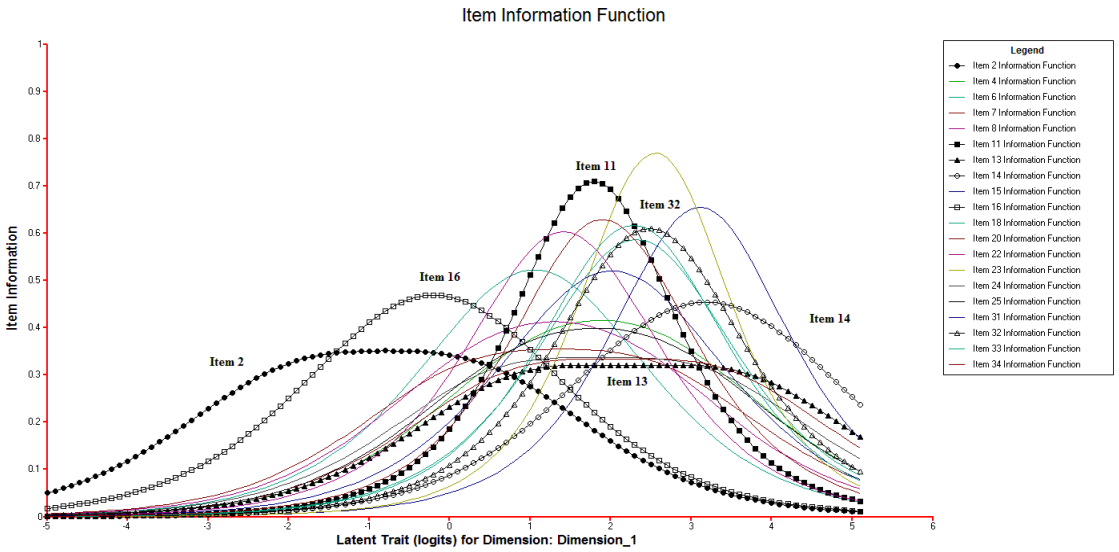

Figure 5. Item information function for Emotion trait (Taiwan).

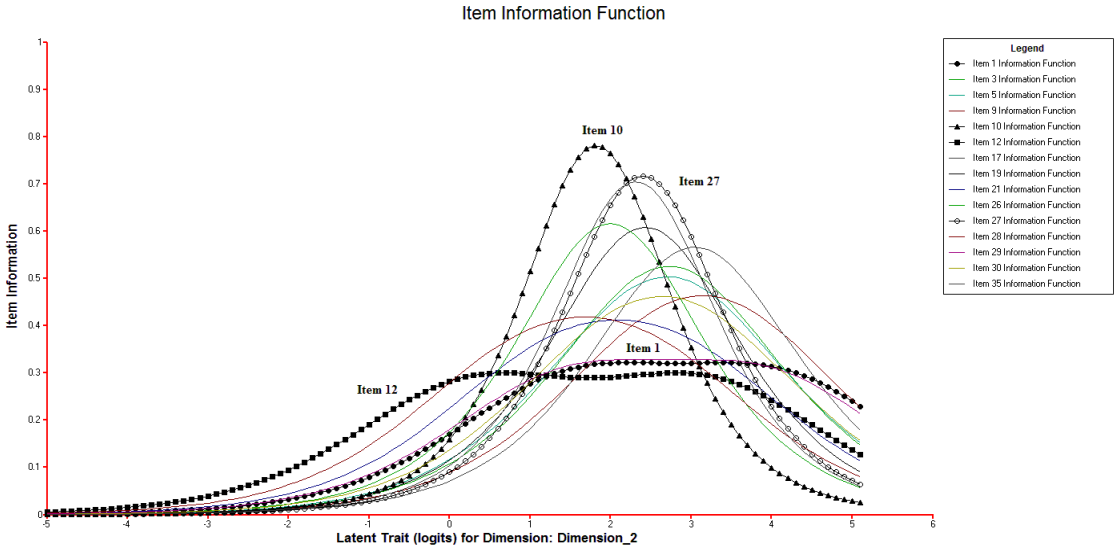

Figure 6. Item information function for Sociality trait (Taiwan).

Differential Item Functioning

The results of DIF analyses indicated that there were 24 items (68.6%) with negligible DIF, five items (14.3%) with slight to moderate DIF, and six items (17.1%) with moderate to large DIF, depending on ETS standards (Zwick, Thayer, & Lewis, 1999). Details for each item are presented in Table 9. The values in Table 9 show the estimates for the country differences in item difficulty.

Table 9. DIF items detected in the 48-month ASQ:SE between the U.S. and Taiwanese sample.

Level A (n = 24, 68.6%) $\|DIF\| \leq 0.43$ logits		Level B (n = 5, 14.3%) 0.64 logits $\geq \|DIF\| \geq$ 0.43 logits		Level C (n = 6, 17.1%) $\|DIF\| \geq 0.64$ logits	
Item	Logits	Item	Logits	Item	Logits
Item 1	-0.32	Item 16	0.49	Item 2	1.82
Item 3	0.27	Item 18	-0.60	Item 12	0.72
Item 4	0.24	Item 20	-0.44	Item 13	-1.06
Item 5	-0.30	Item 21	0.46	Item 17	-0.77
Item 6	-0.15	Item 35	-0.43	Item 19	-0.63
Item 7	-0.11			Item 32	1.28
Item 8	-0.11				
Item 9	0.32				
Item 10	0.12				
Item 11	0.01				
Item 14	-0.09				
Item 15	-0.21				
Item 22	0.39				
Item 23	-0.15				
Item 24	-0.34				
Item 25	0.10				
Item 26	0.16				
Item 27	-0.37				
Item 28	0.29				
Item 29	-0.14				
Item 30	-0.22				
Item 31	-0.24				
Item 33	-0.08				
Item 34	0.12				

Note: DIF level: *positive values* = the item is harder to achieve for Taiwanese sample when it measures an ability, the item is easier to happen for the Taiwanese sample when it measures a problematic behavior; *negative values* = the item is harder to achieve for U.S. sample when it measures an ability, the item is easier to happen for the U.S. sample when it measures a problematic behavior.

Comparison of Theta

The posterior latent ability distribution was calculated for each child by dimension. The U.S. dataset and Taiwanese dataset were combined into one, and then estimated using 2D-RPCM, constraining on cases, generating person's ability (θ) for each child by dimension. Furthermore, the distribution of the person's ability was organized as histograms. It should be noted that these are not necessarily samples that represent the population of the two countries. Findings could be particular to the sample chose in this study. Emotion (Dimension 1) ability is presented in Figure 7, and Sociality ability is presented in Figure 8. The findings by visual comparison indicate the distribution of the posterior latent ability of two datasets was similar to each other, close to the normal distribution. However, a significant number of the U.S. sample located around $\theta = -2$, which did not appear in the Taiwanese sample. This means that for the dataset analyzed here, more students from the U.S. (non-representative) sample measured at the extreme lower bound of the distribution, which in this case means high social (Dimension 2) and emotional (Dimension 1) competences.

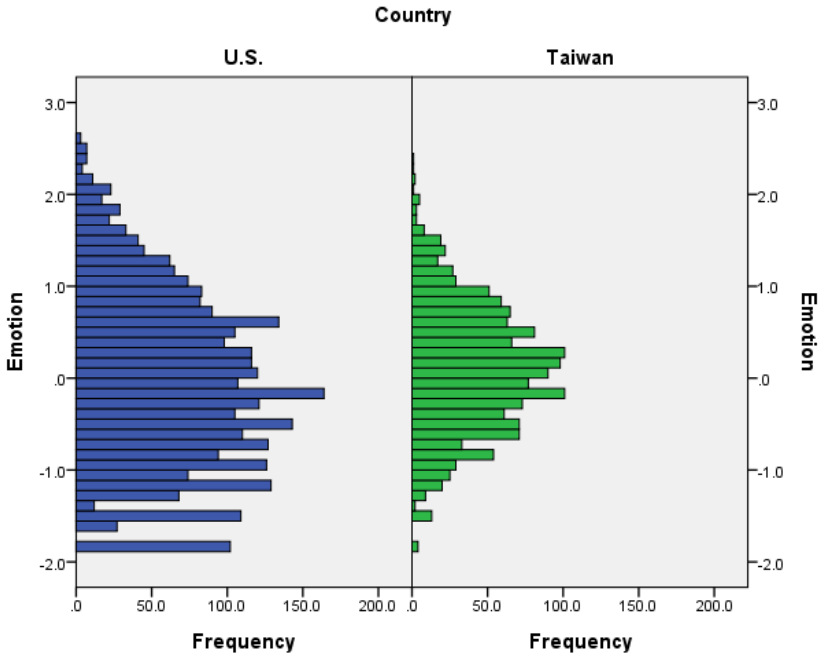

Figure 7. Distribution of latent ability estimates for Emotion trait by country.

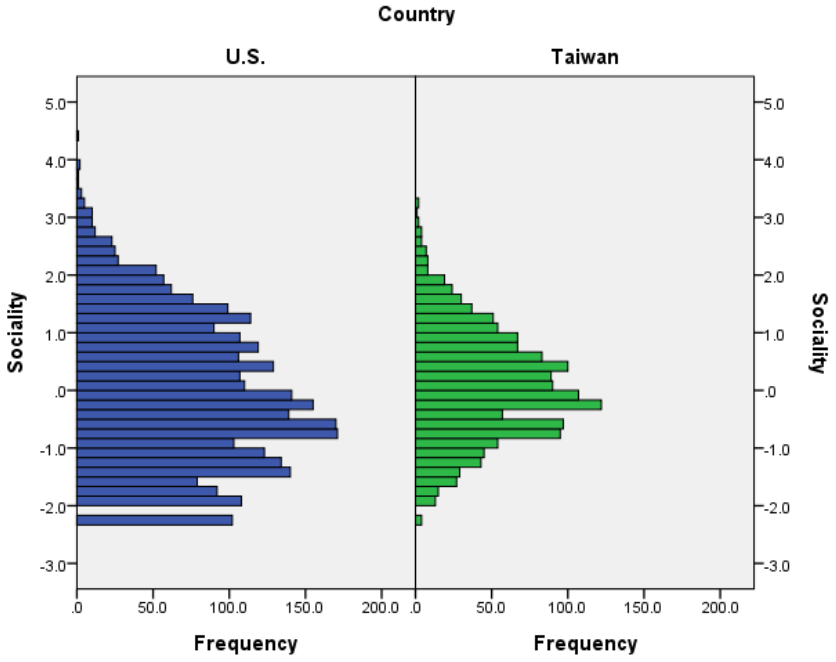

Figure 8. Distribution of latent ability estimates for Sociality trait by country.

CHAPTER V

DISCUSSION

Examining psychometric properties in a translated social-emotional screening test may identify culturally relevant differences between the original version and the translation. This study examined the psychometric properties of the 48-month interval of the Taiwanese Traditional Chinese version of the Ages and Stages Questionnaires: Social-Emotional Second Edition, including investigating its consistency with the original ASQ:SE-2 using differential item functioning, model evaluation, and item fit statistics.

Interpretation of Results

The interpretation of results includes sections related to participants, model evaluation, item fit statistics, item difficulty, reliability, item information functions, DIF, and the comparison of persons' ability.

Participants

The U.S. sample (N = 3,005) was retrieved from an extant dataset collected between 2010 and 2015 as part of a national normative study of the ASQ:SE-2 (Squires et al., 2015). The Taiwanese sample (N = 1,455) was recruited between August 2016 to December 2016 using the ASQ:SE-2 Traditional Chinese version.

The size of the two samples met the minimum size criteria (e.g., at least 500 to 2000) for obtaining accurate parameter estimates using multidimensional IRT (Ackerman, 1994; Jiang, Wang, & Weiss, 2016; Kose & Demirtasli, 2012). For the U.S. sample, the recruitment of families was strategically conducted so that the sample roughly reflected the U.S. 2010 census data on variables including race/ethnicity, mother's education level,

and family income (Squires et al., 2015). The Taiwanese sample was stratified to reflect the 2015 Taiwanese census (Ministry of Health and Welfare, 2016; Ministry of the Interior, 2016). Both samples were collected mostly from parents' responses (i.e., more than 80.0%), and overrepresented parents with higher levels of education. In the U.S. sample, there were 51.0% mothers with education level "college or higher degree," which was 21.0% higher than the 2010 U.S. census; the Taiwanese sample included 14-15% more parents with "college or higher degree" attainment than the census.

Inspecting the Taiwanese sample by collection approach, the paper-pencil sample better reflected the census, with the difference -4.2% to +6.8% in mothers' education, and -2.3% to +0.1% difference in fathers' education compared with the 2015 census (see Table 3). Therefore, findings indicate that parents with higher education seemed to have more access to the Internet than the parents with lower education level in Taiwan, and/or they were more comfortable participating in the study using online questionnaires. Moreover, I found that recruitment through the delivery of paper-pencil copies was more targeted than online collection. Once the kindergarten principal was willing to support this study, families with children at the target age range were recruited.

Model Evaluation

Model evaluation for the unidimensional and multidimensional models was made by comparing model deviance values. The result of the comparison indicated 2D-RPCM fit the U.S. sample and the Taiwanese sample better than the 1D-RPCM did. In addition to suggesting model fit, the results also provided empirical evidence for the internal structure of the ASQ:SE-2 and ASQ:SE-TC, specifically for construct validity of the 48-month interval.

The degree to which the relation between test items and dimensions align with a proposed test construct can be evaluated by examining its internal structure (American Educational Research Association, American Psychological Association, & National Council for Measurement in Education, 2014), which can serve to support construct validity. Based on the theoretical framework of the ASQ:SE-2, social and emotional competence are highly related but represent differing behavioral processes (Squires et al., 2002, 2015). Therefore, a two dimensional model, 2D-RPCM, was posited. The results of dimensional analyses, suggested 2D-RPCM provided adequate evidence to support its theoretical framework. In addition, there were strong correlations between children's posterior latent traits (r = 0.79/0.63), estimated in 2D-RPCM, supporting the hypothesized relations between social competence and emotional competence (Squires et al., 2002; 2015).

In the majority of ASQ:SE validity studies over the last decade across countries, the focus has been on investigating concurrent validity, including sensitivity, specificity, and convergent validity (Alkherainej & Squires, 2016; Heo & Squires, 2012; Jee et al., 2010; Kucuker et al., 2011; Squires et al., 2001; Squires et al., 2002; Squires et al., 2015; Yovanoff & Squires, 2006). However, the evidence for the internal structure of the ASQ:SE has been limited (Gokiert et al., 2014; Kettler & Feeney-Kettler, 2011), except for one study that examined its internal structure using ASQ:SE First Edition (Chen et al., 2016). Thus, these findings add evidence to the existing body of literature regarding the construct validity of the ASQ:SE and its translations.

Item Fit Statistics

The standard applied for evaluating item fit statistics in this study included the range of MNSQ 0.75-1.33 (Wu, Adams, & Wilson, 1998). The results of weighted fit indicated that only one item in the ASQ:SE-2 (1 of 35, 3%) was out of this range, suggesting misfit with the U.S. data. The only misfitting item was Item 6 (MNSQ = 1.43) "Does your child seem too friendly with strangers?" By contrast this item reflected a fit with the Taiwanese dataset (MNSQ = 1.23). With only 3% misfitting items, 2D-RPCM can be considered a good fit status for the U.S. data. These analyses provided evidence that ASQ:SE-2 and ASQ:SE-TC items are able to measure intended underlying constructs.

Item Difficulty

From the information presented in Table 7, the mean item difficulty on ASQ:SE-2 (logits = 2.14) was close to the mean on the ASQ:SE-TC (logits = 2.06). The maximum values of the range were also similar to each other (i.e., U.S. = 3.49 logits; Taiwan = 3.19 logits), but the minimum values of the range presented a lager difference (i.e., U.S. = 0.42 logits; Taiwan = -0.79 logits). In Table 6, Item 2 had minimum values (-0.79 logits) for the ASQ:SE-TC, but its U.S. counterpart presented relatively high logits (1.19 logits). This meant the behavior measured by Item 2 (i.e., Does your child cling to you more than you expect?) occurred more frequently in the Taiwanese sample than in the U.S. sample. However, whether this item functioned differentially between samples needed further tested using DIF analyses.

Referring to the Wright Maps (see Figure 1 and Figure 2), the majority of items were located above $\theta = 1.0$, except for a few items below $\theta = 1.0$ (i.e., Item 2 and Item 16

in ASQ:SE-TC; Item 16 in ASQ:SE-2). The distribution of persons' ability estimates (i.e., the left two panels) described that the majority of children (i.e., the middle of the bell-shaped distribution) was located under $\theta = 1.0$. When a person's ability is at the same point of θ as the item difficulty, the probability of having high scores on the item is 50%. Therefore, most items are relatively difficult for the majority of children in both datasets. This result is consistent with the expectation described in the Methods chapter that the original normative study of ASQ:SE-2 found the majority of children received low scores, which can be explained that items were expected to be difficult for children.

Reliability

The person reliability, so called as EAP/PV reliability, calculated by ConQuest software provides information regarding the replicability of person ordering. The results indicated that the reliability of Emotion (i.e., Dimension 1) and Sociality (i.e., Dimension 2) on the ASQ:SE-2/ASQ:SE-TC was above 0.70, which was recommended for a test (Salvia, Ysseldyke, & Bolt, 2013). Comparing the reliability of the ASQ:SE-2 with ASQ:SE-TC, the reliability of ASQ:SE-2 was higher than ASQ:SE-TC. This result might be explained by Figures 7 and 8, in which the distribution of persons' ability in the U.S. sample had a wider spread than the Taiwanese sample, resulting in demonstrating a hierarchy of ability and an increasing reliability.

Item Information Functions

Items with higher information represent a lower standard error of measurement and higher reliability (DeMars, 2010). Some items on a test can be informative by measuring a certain range of latent traits, whereas some items may not be precise when measuring a certain range of ability. As the dynamic interaction between information (y-

axis) and latent ability (x-axis), the visual analysis has been traditionally used instead of presenting a series of values.

Comparing U.S. item information function curves on the Emotion trait (Figure 3) with the curves for the Taiwanese sample (Figure 5), most items were located in a similar relative position. For example, Item 11 (Does your child have eating problems?) was the most informative item when estimating a person located around $\theta = +1$ to $+2$. Item 16 (Does your child seem more active than other children his age?) was informative when estimating a person with relatively low emotional problems ($\theta < 0$). Item 13 (Does your child do what you ask her to do?), with a flat slope for both samples, was not informative across all continuum of the Emotion trait. Item 14 (Does your child seem happy?) was informative when estimating a person with more emotional problems ($\theta > +3.0$).

Despite the majority of items located in the similar θ positions for both samples, Item 2 (Does your child cling to you more than you expect?) was identified as specifically informative for the Taiwanese sample when located at $\theta < -1.5$, while its peak was $\theta = +1.0$ for the U.S. sample. Item 32 (Does your child show an unusual interest of sexual language and activity?) was identified as extremely informative for the U.S. children, with the Emotion trait located between $\theta = +2.5$ to $+4$, and the summit of the slope reaching 0.85 item information value (y-axis), while its Taiwanese counterpart was informative between $\theta = +2$ to $+3$, with the summit of the slope reaching only 0.60 item information value.

Comparing the U.S. item information function curve for the Sociality trait (Figure 4) with the curves for the Taiwanese sample (Figure 6), most were located in a similar relative position. For example, Item 27 (Can your child name a friend?) and Item 10

(Does your child stay dry during the day?) were informative when estimating a person located around $\theta = +1$ to $+3$; Item 1 (Does your child look at you when you talk to him?), with a flat slope for both samples, was not informative across the ability continuum.

In spite of the majority of items located in similar θ positions for both samples, Item 12 (Do you and your child enjoy mealtimes together?) was identified as specifically informative for the Taiwanese sample with the Sociality trait located in $\theta < 0$. In addition, the shape of the slope of Item 12 (Figure 6) was much flatter than for its U.S. counterpart (Figure 4).

These results provided insight into the quality of the items on the two versions. First, some items functioned significantly differently in the different cultures, such as Item 2, Item 12, and Item 32, which may reflect different parenting or cultural values in the two countries. Second, some items were not informative consistently across samples, such as Item 1 (Does your child look at you when you talk to him?) and Item 13 (Does your child do what you ask her to do?). From a statistics perspective, these results implied that these two items did not measure the construct as precisely as other items (i.e., Item 1 for Sociality; Item 13 for Emotion). When using these items to estimate a person's latent trait, a larger standard error would be identified. However, whether the items should be rewritten or excluded from the ASQ:SE-2 still needs further study as research to date is supportive of item content (e.g., Squires et al., 2002; 20015). For example, Item 1 asking about eye contact when children engage in conversation is a critical indicator for assessing social development. Children who do not have this skill might be a significant concern for parents and professionals (e.g., autism). Although the item function analyses

have provided quantitative suggestions, further studies should be conducted before changes are made.

Differential Item Functioning

DIF analyses investigate whether there are differences between two groups with the same level of a latent trait (Embretson & Reise, 2000). The findings from DIF analysis indicated 14.3% of items (n = 5) with slight to moderate DIF and 17.1% (n = 6) with moderate to large DIF between U.S. and Taiwanese samples.

Level A items may be considered to have negligible DIF. These are more culturally equivalent since parents responded with similar probability across countries. Figure 9 and 10 are examples for Level A items. Figure 9 shows item characteristic curves for Item 3, for U.S. and Taiwan. (The solid curve is for U.S. and the dot curve is for Taiwan.) Given a particular ability level, the probability of getting higher scores on this item is greater for Taiwanese children than for U.S. children. That is, the Taiwanese sample had higher scores (or more problems) on this item than the U.S. sample (i.e., high scores are not desirable). However, the differences between groups were not obvious in Level A items.

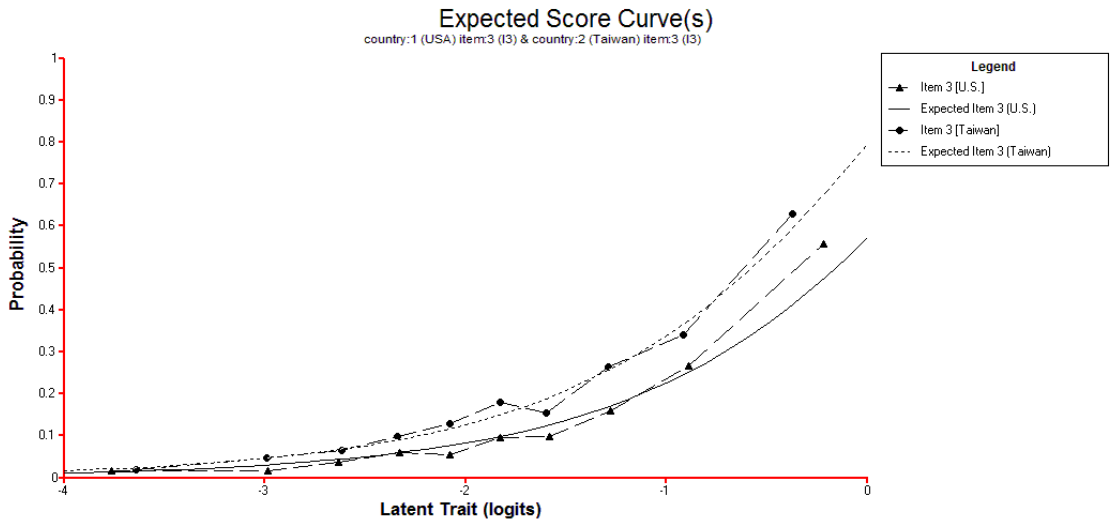

Figure 9. Item characteristic curves for Item 3 (Level A).

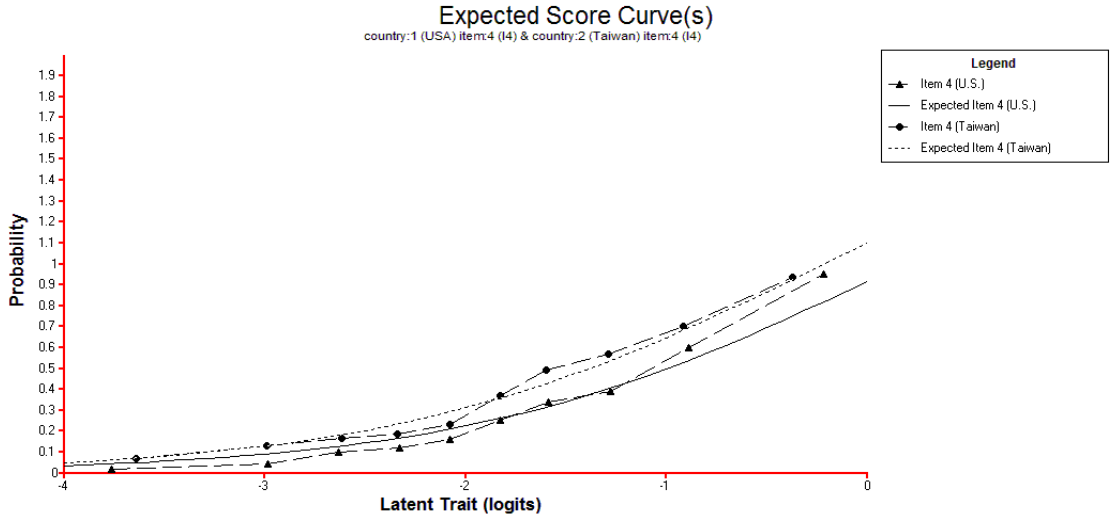

Figure 10. Item characteristic curves for Item 4 (Level A).

Level B items reflect slight to moderate DIF. Figure 11 and Figure 12 are examples of Level B items. The differences indicated in the figures are also not obvious, similar to the figures for Level A, since these items (e.g., Item 20 and Item 35) were just above the critical points of Level B (0.43 logits), which can still be considered a small DIF value.

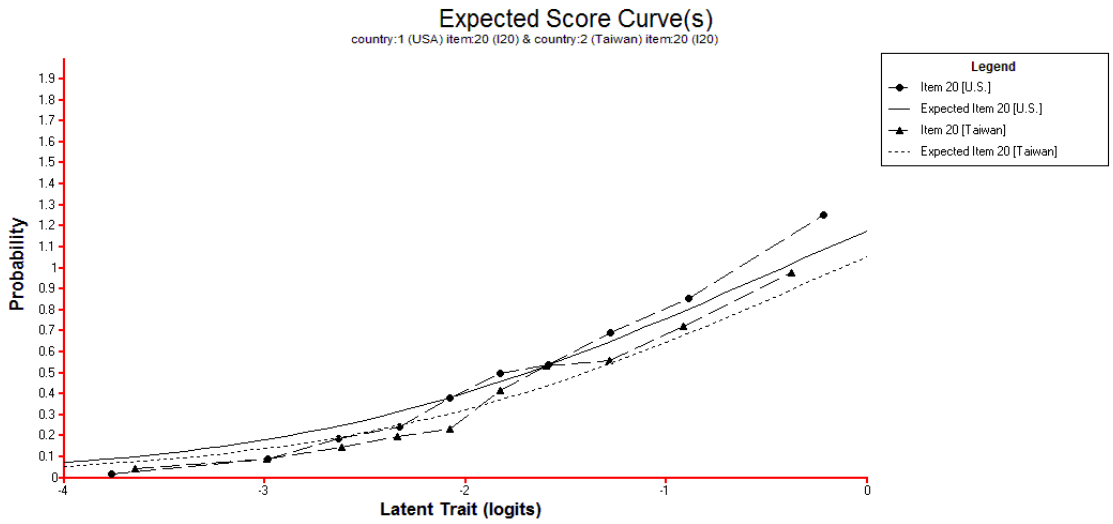

Figure 11. Item characteristic curves for Item 20 (Level B).

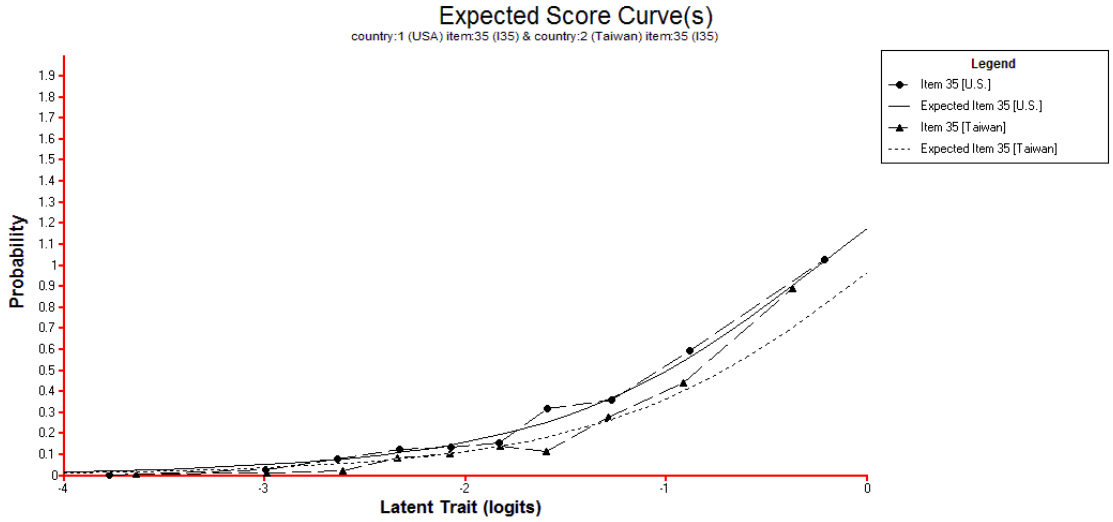

Figure 12. Item characteristic curves for Item 35 (Level B).

Level C items are more problematic since moderate to large DIF might indicate culture differences due to social values, parenting style, or childrearing practices. These differences might result in measurement bias or unfairness. Therefore, exploration of the cultural differences can inform future efforts to translate and adapt the ASQ:SE-2 and other developmental measures in diverse, international settings.

Item 2 (cling more than expect) is the item with largest DIF in the current study. The difference of logits is 1.82, as shown in Figure 13 (solid curve is U.S.; dotted curve Taiwan). Given a particular ability level, the probability of getting higher scores on this item is greater for Taiwanese sample. Compared to Level A and Level B, the differences between groups for Item 2 is substantial. This difference implies that Taiwanese parents had greater probability of considering their children as "clinging to them more than they expect" than did U.S. parents, when two groups of children had the same level of ability. By contrast, Item 13 (-1.06 logits), "Does your child do what you ask her to do?" indicated that U.S. parents had a lower probability to rate their children as doing what they ask him/her to do (see Figure 14).

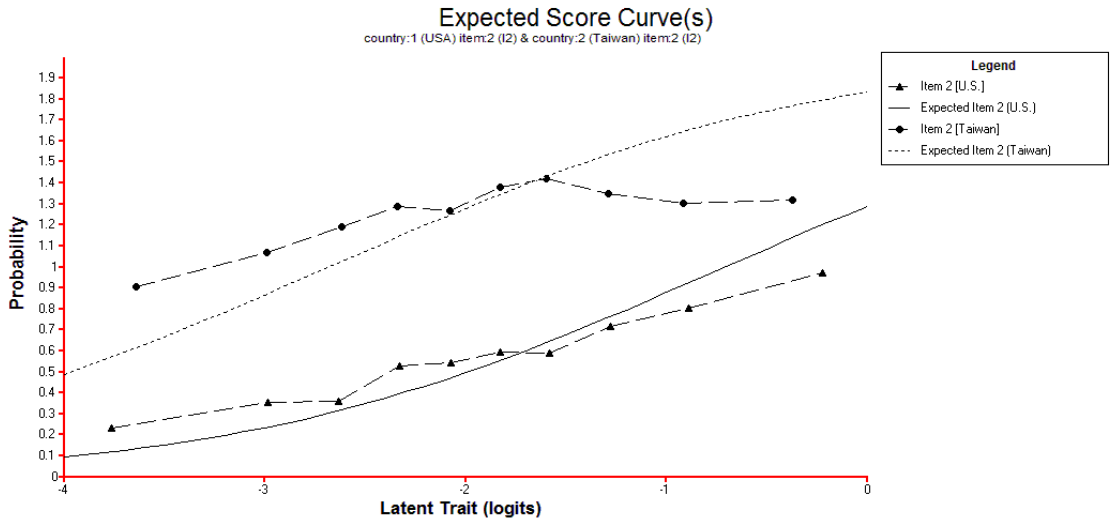

Figure 13. Item characteristic curves for Item 2 (Level C).

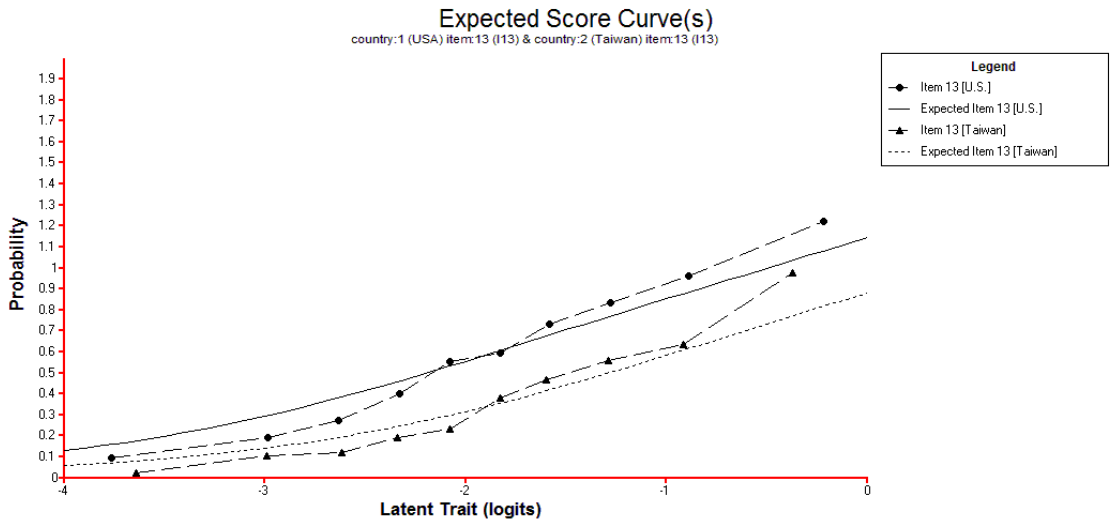

Figure 14. Item characteristic curves for Item 13 (Level C).

Investigating potential cultural reasons for Item 2 (Does your child cling to you more than you expect) and Item 13 (Does your child do what you ask her to do?) may indicate that U.S. children were rated as more likely to demonstrate autonomy or disobedience than were children in Taiwan. Such differences could be explained by the distinction between emphasizing individualism and collectivism as societal values. Chao's (1995) study on mothers' childrearing beliefs pointed out Taiwanese mothers addressed the importance of obedience and respect for parents, whereas European American mothers emphasized the importance of fostering independence. In most Asian cultures, interdependence is a core value, whereas North Americans tend to focus on the individual (Harrison, Wilson, Pine, Chan, & Buriel, 1990; Spence, 1985; Triandis et al., 1986). Thus, childrearing practices in the U.S. often focus on promoting autonomy and consider disobedience less of an issue in young children than in Asian cultures (Kim, Kim, & Rue, 1997; Lin & Fu, 1990; Okagaki & Sternberg, 1993). This rationale can also be applied to explain the Level B DIF detected in Item 20 (Can your child move from one

activity to the next with little difficulty?) and Item 21 (Does your child explore new places, such as a park or a friend's home?).

The current findings regarding Level C items are consistent with Chen et al.'s (2017) cultural comparison between U.S. and Taiwan data on the 60-month interval of ASQ:SE, 1st Edition. In addition to Item 2 and 13, other Level C items (i.e., Item 12, Item 17, Item 19, and Item 32) were the same. They explained the identified DIF in Item 12 (Do you and your child enjoy mealtimes together?) via expert feedback. Generally, U.S. parents considered Item 12 as a common home occurrence. However, in Taiwan, experts felt that mealtimes can be challenging due to cultural beliefs. There is a common saying in Taiwan from *Tang Poems* that "Every single grain is the fruit of hard work" and this has frequently been used to teach children to cherish food. Taiwanese religious tradition also warns that if people have leftovers, they will be punished by staying in hell to eat what they wasted in their lifetime. These traditions have a strong influence on Taiwanese childrearing practices, so that parents often insist that children eat all their food and they are not allowed to be picky eaters, which may result in struggles at mealtimes.

U.S. parents rated fewer concerns than Taiwanese parents on Item 32 (Does your child show an unusual interest in sexual language and activity?), consistent with Chen et al. (2017). U.S. parents reported fewer concerns on this item than parents from East Asian countries including China, South Korea, and Taiwan. Expert consultants suggested that sexuality education and cultural beliefs may have contributed to the DIF results. For example, sexuality education is more accessible in the U.S. than in China, and parents in the U.S. are more open to discussing sexually-related topics with their children (Zhou, 2012). Furthermore, Taiwanese parents seldom talk about sex with children because of

discomfort, embarrassment, traditional social value, and/or lack of sexuality knowledge, so that many parents expect school teachers to have the responsibility to present sexuality education for their children (Lu & Lo, 2014).

Item 17 (Does your child use words to tell you what she wants or needs?) and Item 19 (Does your child use words to describe her feelings and the feelings of others?) are items measuring social-communication skills on the ASQ:SE-2. The DIF results show that Taiwanese parents had a higher probability of rating their children as having a better performance. This finding conflicts with general impressions of American culture (i.e., individualism) and Chinese culture (i.e., collectivism), as well as Chao (1995) regarding differences between American and Taiwanese mothers' childrearing beliefs. Specifically, Chao noted that U.S. parents addressed two beliefs counter to the current finding. First, 40% of European American mothers (20 out of 50) endorsed "processing feeling and emotional honesty", while no Taiwanese mother addressed this topic. American mothers tended to help children be aware of their own feelings and to convey their own emotions so that they can "get their needs met." Second, European American mothers emphasized their child's individuality and self-expression. For example, one mother said she encouraged her child "… to do things on her own and question things…and try to let her come to some of her conclusions." The current findings and Chen et al. (2017) had similar results, with significant DIF preferring the Taiwanese samples. There are several possible explanations.

The first is related to parenting styles. East Asian parents attempt to gratify their infant's early needs immediately, such as picking them up when they cry, and carrying them much of the time (Lynch & Hanson, 2011). Specifically, Chinese parents are often

more protective and controlling than European American parents (Chao, 2000; Chen et al., 1998; Lin & Fu, 1990), and they also often keep their young children physically close to them (Ho, 1986). This parenting style might make the parents highly responsive to their children's need, so that they recognize their children's intentions easily.

The second reason is that since U.S. mothers encourage their children to express their own needs and feelings as one of their childrearing beliefs (Chao, 1995) so they may have higher expectations for these behaviors. By contrast, the Chinese culture encourages people to inhibit behavioral and emotional expressions of individual needs and desires in order to fit in the society (Ho, 1986). The expectation of self-expression therefore would not be as high as it is for U.S. parents. If this postulation is correct, the lower probability of Taiwanese parents rating their children as having these skills could be explained.

In conclusion, given the large portion (n = 6, 17.1%) of Level C DIF items identified between original English and Traditional Chinese versions of the ASQ:SE-2, professionals should be cautious when administering translated assessments. Cultural practices including social values, childrearing, and parenting styles should also be taken into consideration.

Comparison of Theta

By examining Figures 7 and 8, it can be seen that the posterior latent ability distribution of the U.S. and Taiwanese datasets both presented a bell-shaped distribution. The U.S. sample had a wider distribution than the Taiwanese sample, because the size of the U.S. sample is two times bigger than its Taiwanese counterpart. A significant number of the U.S. sample was located around $\theta = -2$, which did not appear in the Taiwanese sample. Reviewing the raw scores in these datasets, the U.S. sample contained 36.2% of

the sample with lower than 25 points for ASQ:SE-2 total scores, while Taiwanese sample contained only 19.1% sample with lower than 25 points. (Each item has scores of 0, 5, or 10, without adding the extra concern scores.) The obvious difference of low score percentages may have resulted in the differences between latent ability estimates around $\theta = -2$ location.

Studies regarding response styles might explain the reason why the U.S. sample had a larger portion of low scores. East Asian populations often prefer to answer using a middle response option when answering a questionnaire (that is, not high or low), whereas Western populations tend to use high or low response options (Chen, Lee, & Stevenson, 1995; Harzing, 2006; Mayer, Elliott, Haas, Hays, & Weinick, 2016; Wang, Hempton, Dugan, & Komives, 2008). Johnson, Kulesa, Cho, and Shavitt (2005) made an assumption that the motive of using an extreme response style is to achieve clarity, precision, and decisiveness in one's explicit statements, while using a middle response style conforms to the standards with norms for ambiguity, flexibility, and modesty in one's statements. The significant number of low score responses in the U.S. sample might reflect the pattern that U.S. parents used the low score options most frequently for insignificant problem behaviors. In contrast, Taiwanese parents might have used "sometimes" to answer an item, even though they did not think their child had a problem. Nevertheless, this explanation needs further exploration.

Limitations of the Study

Several study limitations should be noted including: (1) sample attrition, (2) characteristics of participants, (3) variation in collection process, and (4) implications of psychometric properties.

Sample Attrition

For the Taiwanese sample, some participants dropped out during the recruitment process. A total of 500 hard copies were sent to the kindergartens and 444 copies were returned. The reasons for attrition included: (a) participants disagreed or lost the questionnaires, and (b) several kindergarten principals who consented to participate refused later on due to their busy schedules. For online collection, a total of 1,786 people entered the survey website, but only 1,011 (56.6%) participants completed the questionnaires. Reasons for attrition included: (c) participants changed their mind (n = 8); (d) participants' children were not in the targeted age range (n = 350), (e) participants omitted more than one-third of items (n = 417) were not included in analyses.

Sample attrition possibly led to selection bias. There is a possibility that people who did not complete online questionnaires or omitted most of the items had some specific characteristics. For example, they might not have been familiar with computers, or they might have had a too busy family life, or might have had difficulty in reading Chinese. This selection bias may have resulted in missing data from some subgroups in the Taiwanese society, compromising the representativeness of the data.

Characteristics of Participants

Both U.S. and Taiwanese samples overrepresented parents with college or higher education level, especially for the online samples. The education level for the paper-pencil sample was consistent with respective country census counts. This indicated that parents with higher education levels may have had more access to the Internet, or they might have been more likely to pay attention to a child-related study online. In addition, the number of boys (n = 1,815, 60.4%) was higher than girls (n = 1,190, 39.6%) in the

U.S. data. This might affect results related to boys having more noticeable social-emotional problems such as externalizing difficulties (Chen et al., 2015) so that their parents were more likely to search for online assistance. Therefore, a degree of caution is required when interpreting the results, since limitations exist in the study sample.

Variation in Recruitment Process

Using online data collection has numerous advantages including low cost, saving time for data entry and delivery (Wright, 2005), and more environmentally friendly with no paper used. For the current study, online data provided a unique advantage in that the researcher could access families with children who did not attend kindergartens. In addition, kindergarten classrooms in a big city like Taipei tended to be reluctant to participate because of busy schedules, as well as the lack of relationship with the author. Online collection provided direct access to parents.

Nevertheless, careless or inattentive responding has been noted as a concern for online collection (Johnson, 2005). Johnson (2005) noted that the distance between the administrator and participants may make participants feel less accountable for their questionnaires, and the ease of responding online might make them more careless than when using a paper-pencil copy. With this concern, a further DIF analysis was conducted to examine whether the items functioned differentially between online and paper-pencil versions. The result was acceptable, with 28 items at Level A, indicating 80.0% of the items functioned similarly between the versions. However, there were still six items at Level B (17.1%) and one item at Level C (2.9%) that were not acceptable? Even though the majority of items were equivalent, 20% of the items still arouse certain levels of concern. Caution applying the results of this study is suggested.

Implication of Psychometric Properties

Although the findings present promising psychometric properties for the ASQ:SE-2 and ASQ:SE-TC, including consistency between theoretical frameworks, and internal structure, robust item fit statistics, adequate reliability and item difficulty, and appropriate distribution of personal ability, the design of the study did not allow for the examination of concurrent validity. Thus, establishing cutoff scores for identifying children "at-risk" in Taiwan is an important outcome that still needs to be undertaken.

Future Directions

This study examined the psychometric properties of the 48-month interval of the ASQ:SE-2 and ASQ:SE-TC, as well as differences regarding item functions, item information, and person estimates between the original and translated versions. Social-emotional behaviors are quite different between different periods of childhood (e.g., infants, toddlers, and preschoolers). Cultural differences at 6 months of age may or may not as distinct as they are for children at 4 years. Future studies should focus on the examination and comparison for other ASQ:SE-2 intervals at different ages (i.e., 2-, 6-, 12-, 18-, 24-, 30-, 36-, 60-month interval). In addition, for the practical use of the Traditional Chinese version of the ASQ:SE-2, it will be necessary to focus on concurrent validity for deciding the cut-off points.

Conclusion

Through the model evaluation using unidimensional and multidimensional IRT models, the internal structure of the 48-month interval of the ASQ:SE-2 and ASQ:SE-TC was supported. The theoretical relation between two dimensions, the Sociality and Emotion, was also supported by the results of multidimensional IRT modeling. Adequate

quality of items on the ASQ:SE-2 and ASQ:SE-TC was supported, based on the weighted fit statistics. Only one item was identified as misfitting in the U.S. dataset. Item difficulty was consistent with the expectation that items were difficult for children with typical social-emotional competence. Reliability estimated for a total of 35 items of the ASQ:SE-2/ASQ:SE-TC was 0.86/0.79; for the Emotion dimension was 0.84/0.75; and for Sociality was 0.81/0.74. Item information curves provided information about standard error of each item across the continuum of latent traits. The DIF analyses results showed that there were six items (17.1%) with moderate to large DIF. The comparisons of Theta presented a bell-shaped distribution for both datasets, while some particular differences were identified. Cultural explanations were made for interpreting the potential differences shown in DIF analyses and the comparisons of Theta.

Investigating a social-emotional screener targeting young children may enhance the understanding of early development and assessment. This study contributes to the body of knowledge related to international developmental screening practices and provides supportive evidence for the psychometric properties for the ASQ:SE-2 and its Tradition Chinese translation.

REFERENCES CITED

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.

Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education, 7*(4), 255-278.

Adams, R. J., Wilson, M., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*(1), 1-23.

Adams, R. J., Wu, M. L., & Wilson, M. R. (2016). ACER ConQuest: Generalised Item Response Modelling Software Version 4.5 [computer software]. Camberwell, Australia: ACER.

Alkherainej, K., & Squires, J. (2016). Accuracy of three screening instruments in identifying preschool children at risk for autism spectrum disorder. *Journal of Intellectual Disability-Diagnosis and Treatment, 3*(4), 156-163.

Allen, S. G., Berry, A. D., Brewster, J. A., Chalasani, R. K., & Mack, P. K. (2010). Enhancing developmentally oriented primary care: An Illinois initiative to increase developmental screening in medical homes. *Pediatrics, 126*(Supplement 3), S160-S164.

American Educational Research Association, American Psychological Association, & National Council for Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Ansley, T. N., & Forsyth, R. A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement, 9*(1), 37-48.

Baggett, K. M., Warlen, L., Hamilton, J. L., Roberts, J. L., & Staker, M. (2007). Screening infant mental health indicators: An early head start initiative. *Infants & Young Children, 20*(4), 300-310.

Beeber, L. S., Holditch-Davis, D., Perreira, K., Schwartz, T. A., Lewis, V., Blanchard, H., . . . Goldman, B. D. (2010). Short-term in-home intervention reduces depressive symptoms in Early Head Start Latina mothers of infants and toddlers. *Research in Nursing and Health, 33*(1), 60-76.

Betancourt, H., & Lopez, S. R. (1993). The study of culture, ethnicity, and race in American psychology. *American Psychologist, 48*(6), 629-637.

Bian, X. Y., Wang, R., & Chen, C.-I. (2013). *Ages and Stages Questionnaires: Social Emotional-Simplified Chinese version*. Shanghai, China: Shanghai Scientific and Technical Publishers.

Bond, T., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. New York, NY: Routledge.

Bricker, D., Davis, M. S., & Squires, J. (2004). Mental health screening in young children. *Infants & Young Children, 17*(2), 129-144.

Bricker, D., & Squires, J. (2014). *A history and overview of Ages and Stages Questionnaires and Ages and Stages Questionnaires: Social-Emotional*. Baltimore, MD: Paul H. Brookes Publishing.

Briggs-Gowan, M. J., & Carter, A. S. (2008). Social-emotional screening status in early childhood predicts elementary school outcomes. *Pediatrics, 121*(5), 957-962.

Briggs-Gowan, M. J., Carter, A. S., Irwin, J. R., Wachtel, K., & Cicchetti, D. V. (2004). The Brief Infant-Toddler Social and Emotional Assessment: Screening for social-emotional problems and delays in competence. *Journal of Pediatric Psychology, 29*(2), 143-155.

Briggs-Gowan, M. J., Carter, A. S., Skuban, E. M., & Horwitz, S. M. (2001). Prevalence of social-emotional and behavioral problems in a community sample of 1-and 2-year-old children. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(7), 811-819.

Briggs, R. D., Stettler, E. M., Silver, E. J., Schrag, R. D. A., Nayak, M., Chinitz, S., & Racine, A. D. (2012). Social-emotional screening for infants and toddlers in primary care. *Pediatrics, 129*(2), e377-e384.

Bronfenbrenner, U. (1979). The ecology of human development: Experiments by nature and design: Cambridge, MA: Harvard University Press.

Brown, C. M., Copeland, K. A., Sucharew, H., & Kahn, R. S. (2012). Social-emotional problems in preschool-aged children: Opportunities for prevention and early intervention. *Archives of Pediatrics & Adolescent Medicine, 166*(10), 926-932.

Cai, L., Thissen, D., & du Toit, S. (2011). *IRTPRO user's guide*. Lincolnwood, IL: Scientific Software International.

Campbell, S. B. (1995). Behavior problems in preschool children: a review of recent research. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 36*(1), 113-149.

Campos, J. J., Mumme, D. L., Kermoian, R., & Campos, R. G. (1994). A functionalist perspective on the nature of emotion. *Monographs of the Society for Research in Child Development, 59*(2-3), 284-303.

Carter, A. S., Briggs-Gowan, M. J., Jones, S. M., & Little, T. D. (2003). The Infant-Toddler Social and Emotional Assessment (ITSEA): Factor structure, reliability, and validity. *Journal of Abnormal Child Psychology, 31*(5), 495-514.

Carter, A. S., Briggs-Gowan, M. J., & Davis, N. O. (2004). Assessment of young children's social-emotional development and psychopathology: Recent advances and recommendations for practice. *Journal of Child Psychology and Psychiatry, 45*(1), 109-134.

Chao, R. K. (1995). Chinese and European American cultural models of the self reflected in mothers' childrearing beliefs. *Ethos, 23*(3), 328-354.

Chao, R. K. (2000). The parenting of immigrant Chinese and European American mothers: Relations between parenting styles, socialization goals, and parental practices. *Journal of Applied Developmental Psychology, 21*(2), 233-248.

Chen, C., Lee, S.-Y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, 170-175.

Chen, C. Y., Chen, C.-I., Squires, J., Bian, X. Y., Heo, K. H., Filgueiras, A., . . . Landeira-Fernandez, J. (2017). Adapting a developmental screening measure: Exploring the effects of language and culture on a parent-completed social-emotional screening test. *Infants & Young Children, 30*(2), 111-123.

Chen, C. Y., Filgueiras, A., Squires, J., & Landeira-Fernandez, J. (2016). Examining the factor structure of an early childhood social emotional screening assessment. *The Journal of Special Education and Rehabilitation, 17*(3-4), 89-104.

Chen, C. Y., Squires, J., Heo, K. H., Bian, X. Y., Chen, C.-I., Filgueiras, A., . . . Landeira-Fernandez, J. (2015). Cross cultural gender differences in social-emotional competence of young children: Comparisons with Brazil, China, South Korea, and the United States. *Mental Health in Family Medicine, 11*(2), 59-68.

Chen, X., Hastings, P. D., Rubin, K. H., Chen, H., Cen, G., & Stewart, S. L. (1998). Child-rearing attitudes and behavioral inhibition in Chinese and Canadian toddlers: A cross-cultural study. *Developmental Psychology, 34*(4), 677-686.

Cole, M. (1998). *Cultural psychology: A once and future discipline*. Cambridge, MA: Harvard University Press.

Collins, W. A. (2002). Historical perspectives on contemporary research in social development. In P. K. Smith & C. H. Hart (Eds.), *The Wiley-Blackwell handbook of childhood social development* (pp. 3-24). Malden, MA: Blackwell.

Davis, M. S. (2002). A comparison of three social emotional screening instruments. (Unpublished doctoral dissertation). University of Oregon, Eugene, Oregon.

De Ayala, R. J. (2009). *Theory and practice of item response theory*. New York, NY: Guilford Publications.

DelCarmen-Wiggins, R., & Carter, A. S. (2004). *Handbook of infant, toddler, and preschool mental health assessment*. New York, NY: Oxford University Press.

DeMars, C. (2010). *Item response theory: Understanding statistics measurement.* New York, NY: Oxford University Press.

Denham, S. A. (1998). *Emotional development in young children*. New York, NY: Guilford Press.

Denham, S. A. (2006). Social-emotional competence as support for school readiness: What is it and how do we assess it? *Early Education and Development, 17*(1), 57-89.

Denham, S. A., Wyatt, T. M., Bassett, H. H., Echeverria, D., & Knox, S. S. (2009). Assessing social-emotional development in children from a longitudinal perspective. *Journal of Epidemiology and Community Health, 63*(Suppl 1), i37-i52.

Dunn, D. W., Austin, J. K., & Perkins, S. M. (2009). Prevalence of psychopathology in childhood epilepsy: categorical and dimensional measures. *Developmental Medicine & Child Neurology, 51*(5), 364-372.

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research, 16*(1), 5-18.

Embretson, S. E., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Farver, J. M. (1999). Activity setting analysis: A model for examining the role of culture in development. In A. Göncü (Ed.), *Children's engagement in the world: Sociocultural perspectives* (pp. 99-127). New York, NY: Cambridge University Press.

Feil, E. G., Walker, H. M., & Severson, H. H. (1995). The early screening project for young children with behavior problems. *Journal of Emotional and Behavioral Disorders, 3*(4), 194-202.

Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology, 78*(2), 350-365.

Garces, E., Thomas, D., & Currie, J. (2002). Longer-term effects of Head Start. *The American Economic Review, 92*(4), 999-1012.

Coll, C. G., & Magnuson, K. (2000). Cultural differences as sources of developmental vulnerabilities and resources. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention (2nd ed.)*. (pp. 94-114). New York, NY: Cambridge University Press.

Gardiner, H. W., Mutter, J. D., & Kosmitzki, C. (2008). *Lives across cultures: Cross-cultural human development* (4th ed.). Boston, MA: Allyn and Bacon.

Gauvain, M. (2001). *The social context of cognitive development*. New York, NY: Guilford Press.

Gilkerson, L., & Kopel, C. C. (2005). Relationship-based systems change: Illinois' model for promoting social-emotional development in Part C early intervention. *Infants & Young Children, 18*(4), 349-365.

Gokiert, R. J., Georgis, R., Tremblay, M., Krishnan, V., Vandenberghe, C., & Lee, C. (2014). Evaluating the adequacy of social-emotional measures in early childhood. *Journal of Psychoeducational Assessment, 32*(5), 441-454.

Halfon, N., Regalado, M., Sareen, H., Inkelas, M., Reuland, C. H. P., Glascoe, F. P., & Olson, L. M. (2004). Assessing development in the pediatric office. *Pediatrics, 113*(Supplement 5), 1926-1933.

Hambleton, R. K., & Jones, R. W. (1993). An NCME instructional module on comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12*(3), 38-47.

Hambleton, R. K., & Murray, L. (1983). Some goodness of fit investigations for item response models. In R. K. Hambleton (Eds.), *Applications of item response theory* (pp. 71-94). Vancouver, BC: Educational Research Institute of British Columbia

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer-Nijhoff Publishing.

Harkness, S., & Super, C. M. (1996). *Parents' cultural belief systems: Their origins, expressions, and consequences*. New York, NY: Guilford Press.

Harrison, A. O., Wilson, M. N., Pine, C. J., Chan, S. Q., & Buriel, R. (1990). Family ecologies of ethnic minority children. *Child Development, 61*(2), 347-362.

Harzing, A.-W. (2006). Response styles in cross-national survey research a 26-country study. *International Journal of Cross Cultural Management, 6*(2), 243-266.

Heo, G. (1999). Early identification of social-emotional competence in young children: A study of the Ages and Stages Questionnaires: Social-Emotional. (Unpublished

doctoral dissertation). University of Oregon, Eugene, Oregon.

Heo, K. H., Lee, J. S., & Squires, J. (2012). *Korea - The Ages & Stages Questionnaires: Social-Emotional user's guide.* Seoul, South Korea: Seoul Rehabilitation Center.

Heo, K. H., & Squires, J. (2012). Cultural adaptation of a parent completed social emotional screening instrument for young children: Ages and stages questionnaire-social emotional. *Early Human Development, 88*(3), 151-158.

Hetherington, E. M., Parke, R. D., & Locke, V. O. (1999). *Child psychology: A contemporary viewpoint*. Columbus, OH: McGraw-Hill.

Hillen, T., Gafson, L., Drage, L., & Conlan, L. M. (2012). Assessing the prevalence of mental health disorders and mental health needs among preschool children in care in England. *Infant Mental Health Journal, 33*(4), 411-420.

Ho, D. Y. (1986). Chinese patterns of socialization: A critical review. In M. H. Bond (Ed.), *The psychology of the Chinese people* (pp. 1-37). New York, NY: Oxford University Press.

Jee, S. H., Conn, A. M., Szilagyi, P. G., Blumkin, A., Baldwin, C. D., & Szilagyi, M. A. (2010). Identification of social-emotional problems among young children in foster care. *Journal of Child Psychology and Psychiatry, 51*(12), 1351-1358.

Jellinek, M. S., Murphy, J., Little, M., Pagano, M. E., Comer, D. M., & Kelleher, K. J. (1999). Use of the pediatric symptom checklist to screen for psychosocial problems in pediatric primary care: A national feasibility study. *Archives of Pediatrics & Adolescent Medicine, 153*(3), 254-260.

Jiang, S., Wang, C., & Weiss, D. J. (2016). Sample size requirements for estimation of item parameters in the multidimensional graded response model. *Frontiers in Psychology, 7*, 109.

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality, 39*(1), 103-129.

Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles evidence from 19 countries. *Journal of Cross-Cultural Psychology, 36*(2), 264-277.

Jones, S. M., & Bouffard, S. M. (2012). Social and emotional learning in schools: From programs to strategies. *Social Policy Report, 26*(4), 1-22.

Kelderman, H., & Rijkes, C. P. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika, 59*(2), 149-176.

Keller, H. (2007). *Cultures of infancy*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kettler, R. J., & Feeney-Kettler, K. A. (2011). Screening systems and decision making at the preschool level: Application of a comprehensive validity framework. *Psychology in the Schools, 48*(5), 430-441.

Kim, W. J., Kim, L. I., & Rue, D. S. (1997). Korean American children. In G. Johnson-Powell & J. Yamamoto (Eds.), *Transcultural child development: Psychological assessment and treatment* (pp. 183-207). New York, NY: Wiley.

Kose, I. A., & Demirtasli, N. C. (2012). Comparison of unidimensional and multidimensional models based on item response theory in terms of both variables of test length and sample size. *Procedia-Social and Behavioral Sciences, 46*, 135-140.

Kucuker, S., Kapci, E. G., & Uslu, R. I. (2011). Evaluation of the Turkish version of the Ages and Stages Questionnaires: Social-Emotional in identifying children with social-emotional problems. *Infants and Young Children, 24*(2), 207-220.

Ladd, G. W., Birch, S. H., & Buhs, E. S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development, 70*(6), 1373-1400.

Ladd, G. W., Kochenderfer, B. J., & Coleman, C. C. (1996). Friendship quality as a predictor of young children's early school adjustment. *Child Development, 67*(3), 1103-1118.

Lavigne, J. V., Arend, R., Rosenbaum, D., Binns, H. J., Christoffel, K. K., & Gibbons, R. D. (1998). Psychiatric disorders with onset in the preschool years: II. Correlates and predictors of stable case status. *Journal of the American Academy of Child & Adolescent Psychiatry, 37*(12), 1255-1261.

Lavigne, J. V., Gibbons, R. D., Christoffel, K. K., Arend, R., Rosenbaum, D., Binns, H., . . . Isaacs, C. (1996). Prevalence rates and correlates of psychiatric disorders among preschool children. *Journal of the American Academy of Child and Adolescent Psychiatry, 35*(2), 204-214.

Lin, C. Y. C., & Fu, V. R. (1990). A comparison of child-rearing practices among Chinese, immigrant Chinese, and Caucasian-American parents. *Child Development, 61*(2), 429-433.

Linacre, J. M. (2015). *Winsteps Rasch measurement computer program user's guide.* Beaverton, OR: Winsteps.com.

Lu, C. C., & Lo, M. J. (2014). A study on attitudes toward sexuality education for normal preschoolers and special needs among preschool educators in Taichung City. *Child Care and Education, 4*, 1-27.

Lynch, E. W., & Hanson, M. J. (2011). *Developing cross-cultural competence: A guide for working with young children and their families* (4th ed.). Baltimore, MD: Paul H. Brookes Publishing.

Marks, K., & LaRosa, A. (2012). Understanding developmental-behavioral screening measures. *Pediatrics in Review, 33*(*10*), 448-458.

Masten, A. S., & Coatsworth, J. D. (1995). Competence, resilience, and psychopathology. In D. Cicchetti & D Cohen (Eds.), *Developmental psychopathology: Vol. 2. Risk, disorder, and adaptation* (pp. 715-752). New York, NY: Wiley.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174.

Mathiesen, K. S., & Sanson, A. (2000). Dimensions of early childhood behavior problems: Stability and predictors of change from 18 to 30 months. *Journal of Abnormal Child Psychology, 28*(1), 15-31.

Mayer, L. A., Elliott, M. N., Haas, A., Hays, R. D., & Weinick, R. M. (2016). Less use of extreme response options by Asians to standardized care scenarios may explain some racial/ethnic differences in CAHPS scores. *Medical Care, 54*(1), 38-44.

McKenna, J. J., & Gettler, L. T. (2007). Cultural influences on infant sleep biology and the science that studies it: toward a more inclusive paradigm, part II. In G. Loughlin, J. Carroll, & C. Marcus (Eds.), *Sleep in children: Developmental changes in sleep patterns* (pp. 183-221). New York, NY: Marcel Dekker.

McLean, M. E., Bailey, D. B., & Wolery, M. (1996). *Assessing infants and preschoolers with special needs* (2nd ed.). Englewood Cliffs, NJ: Merrill.

Meiser, T. (1996). Loglinear Rasch models for the analysis of stability and change. *Psychometrika, 61*(4), 629-645.

Merrell, K. W., & Holland, M. L. (1997). Social-emotional behavior of preschool-age children with and without developmental delays. *Research in Developmental Disabilities, 18*(6), 393-405.

Ministry of Health and Welfare. (2016). Monthly report on social welfare statistics. Retrieved from http://www.mohw.gov.tw/CHT/Ministry/Index.aspx

Ministry of the Interior. (2016). Department of Statistics. Retrieved from http://www.moi.gov.tw/stat/

My E Government. (2015). Identification of low and medium-low income family. Retrieved from http://www8.www.gov.tw/policy/welfare/page_01-06.html

Neisworth, J. T., & Bagnato, S. J. (2004). The mismeasure of young children: The authentic assessment alternative. *Infants & Young Children, 17*(3), 198-212.

Noll, R. B., Gartstein, M. A., Vannatta, K., Correll, J., Bukowski, W. M., & Davies, W. H. (1999). Social, emotional, and behavioral functioning of children with cancer. *Pediatrics, 103*(1), 71-78.

Okagaki, L., & Sternberg, R. J. (1993). Parental beliefs and children's school performance. *Child Development, 64*(1), 36-56.

Perez, S. M., & Gauvain, M. (2007). The sociocultural context of transitions in early socioemotional development. In C. A. Brownell & C. B. Kopp (Eds.), *Socioemotional development in the toddler years: Transitions and transformations* (pp. 396-419). New York, NY: Guilford.

Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests.* Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1961). *On general laws and the meaning of measurement in psychology.* Paper presented at the Proceedings of the fourth Berkeley symposium on mathematical statistics and probability.

Raver, C. C., & Zigler, E. F. (1997). Social competence: An untapped dimension in evaluating Head Start's success. *Early Childhood Research Quarterly, 12*(4), 363-385.

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. *Journal of American Medical Association, 285*(18), 2339-2346.

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2002). Age 21 cost-benefit analysis of the Title I Chicago child-parent centers. *Educational Evaluation and Policy Analysis, 24*(4), 267-303.

Robarchek, C. A., & Robarchek, C. J. (1998). Reciprocities and realities: World views. *Aggressive Behavior, 24*, 123-133.

Saarni, C. (1999). *The development of emotional competence*. New York, NY: Guilford Press.

Sand, N., Silverstein, M., Glascoe, F. P., Gupta, V. B., Tonniges, T. P., & O'Connor, K. G. (2005). Pediatricians' reported practices regarding developmental screening: Do guidelines work? Do they help? *Pediatrics, 116*(1), 174-179.

Salvia, J., Ysseldyke, J. E., & Bolt, S. (2013). *Assessment in special and inclusive education* (12th ed.). Belmont, CA: Wadsworth.

Scalise, K. (2012). Crowdsourcing and education with relation to the knowledge economy. In N. Karacapilidis (Ed.), *Web-based and blended educational tools and innovations* (pp. 136-150). Hershey, PA: IGI Global.

Shaw, D. S., Gilliom, M., Ingoldsby, E. M., & Nagin, D. S. (2003). Trajectories leading to school-age conduct problems. *Developmental Psychology, 39*(2), 189-200.

Slentz, K. L., Early, D. M., & McKenna, M. (2008). *A guide to assessment in early childhood, infancy to age eight*. WA: Washington State Office of Superintendent of Public Instruction.

Slonim, M. (1991). *Children, culture, and ethnicity: Evaluating and understanding the impact*. New York, NY: Garland Publication.

Spence, J. T. (1985). Achievement American style: The rewards and costs of individualism. *American Psychologist, 40*(12), 1285-1295.

Squires, J. (2000). Identifying social/emotional and behavioral problems in infants and toddlers. *Infant-Toddler Intervention: The Transdisciplinary Journal, 10*(2), 107-119.

Squires, J., Bricker, D., Heo, K. H., & Twombly, E. (2001). Identification of social-emotional problems in young children using a parent-completed screening measure. *Early Childhood Research Quarterly, 16*(4), 405-419.

Squires, J., Bricker, D., & Twombly, E. (2002). *The Ages & Stages Questionnaires: Social-Emotional.* Baltimore, MD: Paul H. Brookes Publishing.

Squires, J., Bricker, D., & Twombly, E. (2004). Parent-completed screening for social emotional problems in young children: Effects of risk/disability status and gender on performance. *Infant Mental Health Journal, 25*(1), 62-73.

Squires, J., Bricker, D., & Twombly, E. (2015). *Ages & Stages Questionnaires: Social-Emotional Second Edition*. Baltimore, MD: Paul H. Brookes Publishing.

Triandis, H. C., Bontempo, R., Betancourt, H., Bond, M., Leung, K., Brenes, A., . . . Setiadi, B. (1986). The measurement of the etic aspects of individualism and collectivism across cultures. *Australian Journal of Psychology, 38*(3), 257-267.

Trohanis, P. L. (2008). Progress in providing services to young children with special needs and their families: An overview to and update on the implementation of the Individuals with Disabilities Education Act (IDEA). *Journal of Early Intervention, 30*(2), 140-151.

Tudge, J. (1991). Education of young children in the Soviet Union: Current practice in historical perspective. *The Elementary School Journal*, 92(1), 121-133.

Vinden, P. G. (2001). Parenting attitudes and children's understanding of mind: a comparison of Korean American and Anglo-American families. *Cognitive Development, 16*(3), 793-809.

Wang, R., Hempton, B., Dugan, J. P., & Komives, S. R. (2008). Cultural differences: Why do Asians avoid extreme responses? *Survey Practice, 1*(3), 1-7.

Whitcomb, S., & Merrell, K. W. (2013). *Behavioral, social, and emotional assessment of children and adolescents*. New York, NY: Routledge.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.

Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication, 10*(3).

Wu, M. L., Adams, R. J., Wilson, M., & Haldane, S. (2007). *ACER ConQuest version 2.0: Generalised item response modelling software.* Victoria, Australia: ACER.

Wu, M. L., Adams, R. J., & Wilson, M. (1998). *User's manual for ACER ConQuest*. Hawthorn, Australia: ACER.

Yovanoff, P., & Squires, J. (2006). Determining cutoff scores on a developmental screening measure: Use of receiver operating characteristics and item response theory. *Journal of Early Intervention, 29*(1), 48-62.

Zeanah, C. H., Boris, N. W., & Larrieu, J. A. (1997). Infant development and developmental risk: A review of the past 10 years. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*(2), 165-178.

Zero To Three. (2005). *Diagnostic classification of mental health and developmental disorders of infancy and early childhood: Revised edition (DC: 0-3R).* Washington, DC: Zero To Three Press.

Zhou, Y.-F. (2012). The comparative research on sex education for adolescents of China and the US. *US-China Education Review, 2*(4), 408-417.

Zins, J. E., Bloodworth, M. R., Weissberg, R. P., & Walberg, H. J. (2007). The scientific base linking social and emotional learning to school success. *Journal of Educational and Psychological Consultation, 17*(2-3), 191-210.

Zwick, R., Thayer, D. T., & Lewis, C. (1999). An empirical Bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement, 36*(1), 1-28.